

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 21:41:12 ; Search time 1895.15 Seconds
(without alignments)
-951.699 Million cell updates/sec

Title: US-08-988-242..1_COPY_1232_1825
Perfect score: 594
Sequence: 1 CAGGTACAGCTAACGGCTT.....AAACGAGCTCCGGCGATA 594

Scoring table: IDENTITY_NUC
Gapp 10.0 , capext 1.0
Searched: 821193 seqs, -1518192014 residues
Total number of hits satisfying chosen parameters: 11642386
Minimum DB seq length: 0
Maximum DB seq length: 10000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pac: *
6: gb_ph: *
7: gb_Pl1: *
8: gb_Pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_sts: *
14: gb_sy: *
15: gb_un: *
16: gb_vl: *
17: em_fun: *
18: em_hum1: *
19: em_hum2: *
20: em_in: *
21: em_om: *
22: em_or: *
23: em_ov: *
24: em_pat: *
25: em_ph: *
26: em_pl: *
27: em_ro: *
28: em_sts: *
29: em_sy: *
30: em_un: *
31: em_vl: *
32: gb_htg1: *
33: gb_htg2: *
34: gb_in1: *
35: gb_in2: *
36: em_ba1: *
37: em_ba2: *
38: em_hum3: *
39: em_hum4: *
40: gb_pr4: *
41: gb_htg3: *
42: gb_htg4: *
43: gb_htg5: *
44: gb_htg6: *



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	
1	594	100.0	3402	5 AR047920	AR047920 Sequence 1	
2	590.8	99.5	3402	5 AR047920	A48910 Sequence 1	
3	590.8	99.5	3402	35 TCU24..1..0	U24190 Trypanosoma	
4	45	7.6	47852	1 MLY023	AL022022 Mycobacter	
5	44	7.4	2402	1 MURHO	L27277 Micrococcus	
c	6	43.2	7.3	2028	7 SSU13228	AJ13228 Spermatoz
c	7	42.8	7.2	767	34 EIMMAX	M30933 E.tenella
c	8	41.6	7.0	5880	16 HSE4	D14486 Equine herp
c	9	41.2	6.9	2479	12 RN87960	U87960 Rattus norv
c	10	40.4	6.8	477	11 AF020275	AF020275 Homo sapi
c	11	40.4	6.8	3932	10 HSMB01297	AL13103 Homo sapi
c	12	40.4	6.8	3941	12 RNLAR2	X83546 R.norvegicus
c	13	40.4	6.8	69350	1 MTW004	AL009198 Mycobacter
c	14	40.4	6.8	14618	10 AP00556	AP00556 Homo sapi
c	15	40.4	6.8	150036	10 AP00557	AP00557 Homo sapi
c	16	40.4	6.8	157086	10 AP00552	AC007957 Homo sapi
c	17	40.4	6.8	15529	44 AC00757	AC009516 Homo sapi
c	18	40.4	6.8	16937	40 AC009516	AF139019 Cepaea ne
c	19	40	6.7	624	35 CRCGCR1	X17207 Chlamydomon
c	20	39.6	6.7	618	7 AF087653	AF087653 Homo sapi
c	21	39.4	6.6	1442	40 AF087653	L38713 Gallus galli
c	22	39.2	6.6	818	4 CHKPRT001	M28100 Gallus galli
c	23	39.2	6.6	1528	4 CHKPRT001	X95503 M.musculus
c	24	39.2	6.6	2771	12 MMZINCEPR	X95504 M.musculus
c	25	39.2	6.6	3732	12 MMZINCFIP	AB018491 Homo sapi
c	26	39	6.6	259	9 AB018491	AB017820 Homo sapi
c	27	38.8	6.5	720	9 AB017820	X12735 Barley Cab-
c	28	38.8	6.5	1030	7 HYCAB2	X68361 M.fasciata
c	29	38.8	6.5	2858	10 MEAPOAMA	M15764 D.melanogas
c	30	38.4	6.5	150	34 DROFISHA	M22760 Homo sapien
c	31	38.4	6.5	2268	9 HUMCOPTII	X619495 D.melanogas
c	32	38.4	6.5	2722	34 DMNG2	AB010111 Mus sp. m
c	33	38.4	6.5	2837	12 AB010711	AF030576 Acidamano
c	34	38.4	6.5	3015	2 DROFISHA	M23222 D.melanogas
c	35	38.4	6.5	5632	34 DROFISHA	M57388 Slinian herp
c	36	38.4	6.5	7792	34 DROFISHA	AF126748 Homo sapi
c	37	38.4	6.5	11196	43 AC013031	AF126749 Homo sapi
c	38	38.4	6.5	37518	43 AC014003	Y08701 M.musculus
c	39	38.4	6.5	39565	34 DMC96610	M33782 Human TFE8
c	40	38.2	6.4	2943	5 I17281	
c	41	38.2	6.4	2943	16 HSVGBQ	
c	42	38	6.4	1159	40 AF126748	
c	43	38	6.4	1472	40 AF126749	
c	44	38	6.4	2813	12 MMPININ	
c	45	37.8	6.4	1526	9 HOMTFEB	

ALIGNMENTS

RESULT	1	AR047920	AR047920	AR047920	AR047920
LOCUS		DEFINITION	DEFINITION	DEFINITION	DEFINITION
ACCESSION		ACCESSION	ACCESSION	ACCESSION	ACCESSION
VERSION		VERSION	VERSION	VERSION	VERSION
KEYWORDS		KEYWORDS	KEYWORDS	KEYWORDS	KEYWORDS
RESULTS	1	AR047920	AR047920	AR047920	AR047920
ALIGMENTS					
PAIR					
DATE					

29-SEP-1999

SOURCE	Unknown.	TITLE	NOVEL TRYpanosoma cruzi ANTIGEN, AND GENE CODING THEREFOR; THEIR APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
ORGANISM	Unclassified.	JOURNAL	BIO MERIEUX (FRA)
REFERENCE	1 (bases 1 to 3402) Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M. Trypanosoma cruzi antigen, gene encoding thereof and methods of detecting and treating chagas disease	COMMENT	Other publication CA 2173957 960222 Other publication AU 316195 960307 Other publication FR 2723589 960216.
AUTHORS		FEATURES	Location/Qualifiers
TITLE		source	1. /strain="G" /organism="trypanosoma cruzi"
JOURNAL	Patent: US 5820864-A 1 13-OCT-1998;	source	/dev_stage="EPIMASTICOTE"
FEATURES	Location/Qualifiers	source	/db_xref="Taxon:5693" /dev_stage="EPIMASTICOTE"
source	1. /organism="unknown"	BASE COUNT	889 a
BASE COUNT	888 a	ORIGIN	818 C 958 g 737 t
ORIGIN		ORIGIN	958 g 737 t
Query Match	100.0%; Score 594; DB 5; Length 3402;	Query Match	99.5%; Score 590.8; DB 5; Length 3402;
Best Local Similarity	100.0%; Pred. No. 4 4e-117;	Best Local Similarity	99.5%; Pred. No. 2.1e-116;
Matches 594;	0; Mismatches 0;	Matches 592;	Mismatches 2;
Indels 0;	Gaps 0;	Indels 0;	Gaps 0;
Qy 1 CAGGTACACCGTAACGGCTTTGCTCAATCGTAAACAGCCGACGGTAGCTGCCCTCGCTGCT 60	Qy 1 CAGGTACACCGTAACGGCTTTGCTCAATCGTAAACAGCCGACGGTAGCTGCCCTCGCTGCT 60		
Db 1232 CAGGTACACCGTAACGGCTTTGCTCAATCGTAAACAGCCGACGGTAGCTGCCCTCGCT 1291	Db 1232 CAGGTACACCGTAACGGCTTTGCTCAATCGTAAACAGCCGACGGTAGCTGCCCTCGCT 1291		
Qy 61 GACATGTCATTGATTGACGATCTCCATCTCGGTCTCCCGCAGGAAGAACAGCAG 120	Qy 61 GACATGTCATTGACGATCTCCATCTCGGTCTCCCGCAGGAAGAACAGCAG 120		
Db 1292 GACATGTCATTGACGATCTCCATCTCGGTCTCCCGCAGGAAGAACAGCAG 1351	Db 1292 GACATGTCATTGACGATCTCCATCTCGGTCTCCCGCAGGAAGAACAGCAG 1351		
Qy 121 CAGGCCAAAACATCGTAGTGCGAGGAAACCGGGTGCTCCCTGGCACT 180	Qy 121 CCAGGCCAAAACATCGTAGTGCGAGGAAACCGGGTGCTCCCTGGCACT 180		
Db 1352 CAGGCCAAAACATCGTAGTGCGAGGAAACCGGGTGCTCCCTGGCACT 1411	Db 1352 CCAGGCCAAAACATCGTAGTGCGAGGAAACCGGGTGCTCCCTGGCACT 1411		
Qy 181 GACCGGGGAGTAGCAGTACAATACGACTCTCGCATCCCTGCATCA 240	Qy 181 GACCGGGGAGTAGCAGTACAATACGACTCTCGCATCCCTGCATCA 240		
Db 1412 GACCGGGCAGTACAGTCATACGATCTCGCATCCCTGCATCA 1471	Db 1412 GACCGGGCAGTACAGTCATACGATCTCGCATCCCTGCATCA 1471		
Qy 241 CCCCTGTTCAAGGCCAGCCAAGGCCAGCCCTCTGCCGGCCGATCGGTGAG 300	Qy 241 CCCCTGTTCAAGGCCAGCCAAGGCCAGCCCTCTGCCGGCCGATCGGTGAG 300		
Db 1472 CCCCTGTTCAAGGCCAGCCAAGGCCAGCCCTCTGCCGGCCGATCGGTGAG 1531	Db 1472 CCCCTGTTCAAGGCCAGCCAAGGCCAGCCCTCTGCCGGCCGATCGGTGAG 1531		
Qy 301 CGCACGTGGAGCAAGTCAATGCTTAATCTAGTAACTGAGTGGGATTAAATGTCACC 360	Qy 301 CGCACGTGGAGCAAGTCAATGCTTAATCTAGTAACTGAGTGGGATTAAATGTCACC 360		
Db 1532 CGCACGTGGAGCAAGTCAATGCTTAATCTAGTAACTGAGTGGGATTAAATGTCACC 1591	Db 1532 CGCACGTGGAGCAAGTCAATGCTTAATCTAGTAACTGAGTGGGATTAAATGTCACC 1591		
Qy 361 CAAAGGAGGGTGTCAAGCACTGGAGGCGGCCAGCAGGCTTAACGGCGGTGACGCC 420	Qy 361 CAAAGGAGGGTGTCAAGCACTGGAGGCGGCCAGCAGGCTTAACGGCGGTGACGCC 420		
Db 1592 CAAGGAGGGTGTCAAGCACTGGAGGCGGCCAGCAGGCTTAACGGCGGTGACGCC 1651	Db 1592 CAAGGAGGGTGTCAAGCACTGGAGGCGGCCAGCAGGCTTAACGGCGGTGACGCC 1651		
Qy 421 ACGACTACGCCCGCACGCCAACAGTCCATACGGCCACAATGGCCGACCTGTGACGGCT 480	Qy 421 ACGACTACGCCCGCACGCCAACAGTCCATACGGCCACAATGGCCGACCTGTGACGGCT 480		
Db 1652 ACGACTACGCCCGCACGCCAACAGTCCATACGGCCACAATGGCCGACCTGTGACGGCT 1711	Db 1652 ACGACTACGCCCGCACGCCAACAGTCCATACGGCCACAATGGCCGACCTGTGACGGCT 1711		
Qy 481 GGATTGGTGGCAGCTTAATGTTGCTCAAGCCGGCTGTCCTCCACAGCCGGGCAA 540	Qy 481 GGATTGGTGGCAGCTTAATGTTGCTCAAGCCGGCTGTCCTCCACAGCCGGGCAA 540		
Db 1712 GGATTGGTGGCAGCTTAATGTTGCTCAAGCCGGCTGTCCTCCACAGCCGGGCAA 1771	Db 1712 GGATTGGTGGCAGCTTAATGTTGCTCAAGCCGGCTGTCCTCCACAGCCGGGCAA 1771		
Qy 541 CAAACAGGAGAAGAAAAAGCCTTCGGGATGTGAAACAGGCTCGTGGCATA 594	Qy 541 CAAACAGGAGAAGAAAAAGCCTTCGGGATGTGAAACAGGCTCGTGGCATA 594		
Db 1772 CCAACAGGAGAAGAAAAAGCCTTCGGGATGTGAAACAGGCTCGTGGCATA 1825	Db 1772 CCAACAGGAGAAGAAAAAGCCTTCGGGATGTGAAACAGGCTCGTGGCATA 1825		
RESULT	2	RESULT	3
A48910	A48910 3402 bp DNA	TCU24190	TCU24190 3402 bp mRNA
DEFINITION	Sequence 1 from Patent WO9605312.	LOCUS	trypanosoma cruzi Tc40 antigen (Tc40) mRNA, complete cds.
ACCESSION	A48910	DEFINITION	trypanosoma cruzi.
VERSION	A48910.1 GI:2302570	ACCESSION	Eukaryota: Euglenozoa; Kinetoplastida; Trypanosomatida;
KEYWORDS		VERSION	Trypanosoma cruzi.
ORGANISM	Trypanosoma cruzi.	KEYWORDS	Eukaryota: Euglenozoa; Schizotrypanum.
REFERENCE	1 (bases 1 to 3402) Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.	SOURCE	Trypanosoma cruzi.
AUTHORS		ORGANISM	Eukaryota: Euglenozoa; Kineto-trypanum.

REFERENCE	1 (bases 1 to 3402)	Db	1472	CCCCCTGTTTACGGCCAGCGAAGGGAGCCGCCATGGCTCTGGCGGCCATGGCTAG 1531
AUTHORS	Lesenechal,M., Duret,L., Cano,M.I., Mortara,R.A., Jolivet,M., Camargo,M.E., da Silveira,J.F. and Paranhos-Baccala,G.	Qy	301	CGGCACGTGGAGAGATCATTTGCTAACTTAGCTAGATGATTAATCTAGCTGGGATTAATTCACC 360
TITLE	Cloning and characterization of a gene encoding a novel immunodominant antigen of Trypanosoma cruzi	Db	1532	CCGCACGTGGGAGCATGGTAACTCTAGTGAATCTAGCTGGGATTAATTCACC 1591
JOURNAL	Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)	Qy	361	CAAAGGAGCAGTCAGACTGGGAGCCACCAAGAGGTACAGGGTGAACGGTCC 420
MEDLINE	97391123	Db	1592	CAAGGAGCAGTCAGACTGGGAGCCACCAAGAGGTACAGGGTGAACGGTCC 420
REFERENCE	2 (bases 1 to 3402)	Db	1592	CAAGGAGCAGTCAGACTGGGAGCCACCAAGAGGTACAGGGTGAACGGTCC 420
AUTHORS	Lesenechal,M., Franco Da Silveira,J., Mortara,R.A., Duret,L., Camargo,M.E., Jolivet,M. and Paranhos-Baccala,G.	Qy	421	ACGACTACCGCCGCAACTACAGGGCACATGGCTAGCTGGGACCTTGAGCGCT 480
TITLE	Submitted (06-APR-1995) Mylene Lesenechal, Biomerieux, 22 rue Saint Jean de Dieu, Lyon 69007, France	Db	1652	ACGACTACCGCCGCAACTACAGGGCACATGGCTAGCTGGGACCTTGAGCGCT 480
JOURNAL	Location/Qualifiers	Qy	481	GGATTTGGGAGCTATACTTGTCGAGCAACTCCATACAGGGCACATGGCTAGCTGGGACCTTGAGCGCT 540
FEATURES	1..3402 /organism="Trypanosoma cruzi"	Db	1712	GGATTTGGGAGCTATACTTGTCGAGCAACTCCATACAGGGCACATGGCTAGCTGGGACCTTGAGCGCT 1771
Source	/strain="G"	Qy	541	CCAACAGGAGAGAAAGGGCTCGGGCAATGTGAAAGAACGACTCCCTGGCATA 594
Direct Submission	/db_xref="GI:5693"	Db	1772	CCAAACAGGAGAGAAAGGGCTCCGGCTGGCTGTGACGAGCTCGCTGGCATA 1825
Submitted (06-APR-1995) Mylene Lesenechal, Biomerieux, 22 rue Saint Jean de Dieu, Lyon 69007, France	/dev_stage="epimastigote"	RESULT	4	
	1..35	MTVO23		MTVO23 47852 bp DNA
	/gene="Tc40"	LOCUS		MTb tuberculosis H37RV complete genome, segment 148/162.
	/gene="Tc40"	DEFINITION		
CDS	/codon_Start=1	ACCESSION	AL022022 AL123456	
	/product="Tc40 antigen"	VERSION	AL022022.1 G1:3261554	
	/protein_id="AAC47657.1"	KEYWORDS		
	/db_xref="GI:7906467"	ORGANISM		
	/translation="MIVTVDLEHNHAKPSNEGRWVSDATEATFNEVEPAQVLAQDSQFY	Mycobacterium tuberculosis		
	LAYTMKRRHIVRVRVSNLKGTVRAHSKTHAVEFVNNSVNAASAGREFFWVNT	Actinomycetales; Actinobacteria; Actinomycetidae;		
	DETEASNGKPDIAARLTVKYXFEKLQDPVTPCSEFINSPQRDPLVYETQANLID	Mycoabacteriaceae;		
	SSSLIERFDVSELEFLQRNITLRLIQVSEVKAHLIDDATEVSVLYAASHGTYQVNLITGAEPNLRK	Mycoabacterium		
	FVIDGSIVANESSRETFAVFTDRKOLALVNWNHSPINETKHYMPCOVQDNGECFNRT	Coles,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,		
	ADGSCYLAIDSNRLEFLHRSRREEQPOPKTSVATAPGCVSSGTDAASSSTINT	Harris,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,		
	TSAAAASPSPVVAFAKAAPPAAKSPLPHTVAGVLAANVNLQGLINTVQSYVSTG	Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,		
	APATTSTTATVSTTTAPOSPTSPYINGRPVAGVLAANVNSASAASPTAAKPKSEK	Hamble,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,		
	SRAVKKLPPDTTRDHEOLLNIGEVKVAKHLLDATEVSVLYAASHGTYQVNLITGAEPNLRK	Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajadhean,M.A., Roberts,J., Rutter,S., Seeger,K., Skelton,S.,		
	ASACETSSVATINAPLHNASLPOAIPGVLAARYQSGEVIEOSLELESVTNT	Squires,R., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S., and		
	SRVLKSLRNPTRTKGVGRGYNAIIMLDHEVRAHISGNRLPOTRDKNIKSRLDEALKEST	Barrell,B.G.		
	LIADSLSRNTPTRTKGVGRGYNAIIMLDHEVRAHISGNRLPOTRDKNIKSRLDEALKEST	Deciphering the biology of Mycobacterium tuberculosis from the		
	TOFTAQLQTIVENVKRELAVLGINGSLSLVKENASLOKELNSIMSGVLDERR	complete genome sequence		
	MREELCTLRSEVAKRATMPDSSRHTATSSRGPSSAPETLALMSVREOYRCORGE	Nature 393 (6685), 537-544 (1998)		
	YMLMACOPSPJLREFSLTRENNAYSELENVENPNDVCSVLLQIETATEKE	ERRatum: [[published erratum appears in Nature 1998 Nov		
	S"	12:306(6707):1901]		
BASE COUNT	889 a 818 c 958 g 737 t	REFERENCE	2 (bases 1 to 47852)	
ORIGIN		AUTHORS	Parkhill,J.	
		JOURNAL	Direct Submission	
		COMMENT	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk	
		NOTES	On Jun 27, 1998 this sequence version replaced gi:2924430.	
		DETAILS	Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web	
		URL	(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.	
		REMARK	Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.	
		REMARK	CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome	
Query Match	99.5%	Score	590.8; DB 35; Length 3402;	
Best Local Similarity	99.7%	Pred. No.	2..1e-16; Indels 0; Gaps 0;	
Matches	592	Conservative		
Qy	1	CAGTACAGCTAACGGTTTGTCTCAATGATCACAGCGACGTAGCTGGCTCTGGCT 60		
Db	1232	CAGTACAGCTAACGGTTTGTCTCAATGATCACAGCGACGTAGCTGGCTCTGGCT 1291		
Qy	61	GACATGTCGATTCGATGAGCATCTCCATCTCGGTCCTCCGAGGGAAACAGCAG 120		
Db	1292	GACATGTCGATTCGATGAGCATCTCCATCTCGGTCCTCCGAGGGAAACAGCAG 1351		
Qy	121	CCAGGCCAAAACATCGTAGTGGCAGCGAACCGGGTGTCTCGGGCACT 180		
Db	1352	CCAGGCCAAAACATCGTAGTGGCAGCGAACCGGGTGTCTCGGGCACT 1411		
Qy	181	GACGGGGAGTAGCTGATGAGCATACAAATGACGTCTGGCTGCTGATCCCTGCA 240		
Db	1412	GACGGGGAGTAGCTGATGAGCATACAAATGACGTCTGGCTGCTGATCCCTGCA 1471		
Qy	241	CCCCCTGTTTACGGCCAGCAAGGCAGCCGGCCTCTGGGGGGGATCGGGCTGAG 300		

binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES	source	Location/Qualifiers	misc_feature
	1. . 47852	/organism="Mycobacterium tuberculosis" /strain="H37Rv" /db_xref="taxon:1773"	gene
	<1. 18	/organism="Mycobacterium tuberculosis" /strain="H37Rv" /db_xref="taxon:1773"	CDS
		/clone="Y13E12"	
		/gene="Rv3494c"	
		/note="Rv3494c, (MTV023_01c), len: 564. Unknown Pro-rich protein similar to several Mycobacterium tuberculosis proteins e.g. MTCT28_14 (515 aa), MTCT19H5_28C (516 aa) and (MTV051_09). Has hydrophobic stretch, possibly signal peptide at N-terminus. FASTA scores: Z970501 MTCT28_14 (515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 34.3% identity in 559 aa overlap; Z97182 MTCT19H5_28 (516 aa) opt: 979 z-score: 567.7 E(): 4.1e-24; 33.5% identity in 555 aa overlap. TParse score is 0.897"	
		/codon_start=1 /trans_table=11 /product="hypothetical protein Rv3494c"	
		/protein_id="CAA1773_1."	
		/db_xref="GI: 2924431"	
		/note="SPREML: 053539"	
		/translation="MIDLAKLQKLSITFAVITYTIVTISYMAIFYLRLPATFGIGTYGVSADFAAGGGDKKNAVNATYRGAVRVESSVGLNPVGTAHMRLSGTTRPVAIAGEQYDLYPENPSTKLQRGFRTRORIGONVADLIRQAELLGSDGTRRELLHEPATNGAGCPYLARLIESARLLYDEANANYPVQLDQGPFLQAIRAGSDIKSLADGLARFTQQLRAADPRLRDLTADAGSRPSPALASLANIGRVCVYHESIEQOLVVFFPEALFAAITSAFASSGRSPSPALASLANIGRLVSPADESREPDMCKTRNDPSTYGRARNIPQEFGPKRAFTVQLRDPRTSYPVGTNPQRGPPIPYGTEVTGIGNILPPNKPFITPPGPPQVACGPRAFHQAPQAPAPPPNDNGPPPTSTWMPKGYPPPPGPPPTGPPPGPACGPQPAQSGPATVYDOLSGAFADPAGTTGTFAPGMGTGASSAENVVDMRDPLQ"	
		/complement(1..1695)	
		<1. . 47852	
		/note="Fragment designated v023. Does not represent a physical clone"	
		/complement(1703..1707)	
		/note="possible RBS for Rv3494c"	
		/gene="lprN"	
		/complement(1706..2860)	
		/gene="lprN"	
		/note="Rv3495c, (MTV023_02c), len: 384. lprN, Similar to Mycobacterium tuberculosis proteins MTCT28_13 (390 aa) and MTCT19H5_29 (402 aa) and (MTV051_08). Probably lipoprotein contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. FASTA scores: GP Z970501 MTCT28_13 (390 aa) opt: 653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and Z97182 MTCT19H5_29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. TParse score is 0.897"	
		/codon_start=1 /trans_table=11 /product="lprN"	
		/protein_id="CAA1773_2."	
		/db_xref="GI: 2924432"	
		/note="SPREML: 053540"	
		/translation="MNTIWLRRIILPASSALLAGCGFGGLNSLPLPGTAGHGEAYSVTTEMADATLPONSPMVMDDTYTGSTAGIVAVORPDSFYAVKLDKNNVLPANAVAKVSQTSLLGSILVLLAPTRDRPCTLVDSRRIETRPTTEEVFSALGVVNNKGVNGALEBTTIDETHOAVAGRSQQFVNLYVPRDIAETLAGRNHRDHIDIDALDGLNRVSEAI LARDKDNIGRALDTLPDAVRLVNQNDRDHITDAAFKLRTAVTSHVLAETKVDFGEDL	
		/complement(1706..2860)	
		/gene="lprN"	
		/complement(5272..6324)	
		/note="Rv3498c"	
		/db_xref="GI: 2924434"	
		/note="SPREML: 053542"	
		/translation="MURKPSKHEDPLRTGIGFLVLCVVLIAFGYSSCLPFWQGKTYDAYFTDAGGITPNSVYSGVKGAVASVLAGNAAKTVTSYDLSIVYQGDOSLAALNVFTQALHDDATPOVRLGAVDLTSRLRNDRDEAGLQLLAHAKSYTIVSLEAEQVINKLVEDQNLPAFDARAALSALESIDDDYAAQISCFVADRKERCPALSRLNLVIANLNERRDXTTEALKRKPITYATLGEVYVGSGPENVNVYSVLPGPLYATVFDLVFQPSKLPPDSLADYLRFQFQERNVILIRKPSP"	
		/complement(5272..6324)	
		/gene="Rv3498c"	
		/complement(5272..6324)	
		/note="Rv3498c, (MTV023_05c), len: 350. Unknown but similar to Mycobacterium tuberculosis proteins MTCT28_10 (346 aa), MTCT19H5_32 (346 aa) and MTV051_05. Hydrophobic region at N-terminus. FASTA scores: Z970501 MTCT28_10 (346 aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in 340 aa overlap; and Z97182 MTCT19H5_30 (508 aa) opt: 821 z-score: 697.0 E(): 2.6e-31; 35.1% identity in 453 aa overlap. TParse score is 0.891"	
		/codon_start=1 /trans_table=11 /product="hypothetical protein Rv3496c"	
		/protein_id="CAA1773_1."	
		/db_xref="SPREML:OS 3541"	
		/translation="NMGRVANLTGSGLRATVIALVAVLGGVTVLSSGNKRITGVYFTAIVGLYPEDQVRYGVYPGEIDMEPEPDSVATMSVKDVVPDVQAVINSPNGLPLGAANINQAAADTLQNGDSUHNLRELQAVGRGDSRGDIEFTVKNLQVLDLSEDEQITQFAGHVASYSQLADDSANIDQTGTLNQALSDLRGFLENNSLNTYQNLNDFAQTLSDQSENTEQVNLHYVAGPNTINVNQYDQGPTINGLSPNANVQFCIGCSFDTAANGSADSYDREICERBLRGPVLRRTVNTYQPMIFPLNTATYQGQIYDTPATEAKSETVPETLWPAAGGSPAPVGNPADLQSLIYVPPAGPAPAPPAGPGERGGC"	
		/complement(2867..2871)	
		/gene="Rv3496c"	
		/note="possible RBS for Rv3495c"	
		/complement(4209..5282)	
		/gene="Rv3497c"	
		/complement(4209..5282)	
		/gene="Rv3497c"	
		/complement(481..481)	
		/note="Rv3497c, (MTV023_04c), len: 357. Unknown but similar to Mycobacterium tuberculosis proteins MTCT19H5_31 (481 aa), MTCT28_11, (515 aa) and MTV051_06. Hydrophobic region at N-terminus. FASTA scores: Z97182 MTCT19H5_31 (481 aa) opt: 611 z-score: 667.7 E(): 1.1e-29; 32.5% identity in 332aa overlap; and Z97050 MTCT28_11 (515 aa) opt: 587 z-score: 641.2 E(): 3.3e-38; 30.1% identity in 335 aa overlap. TParse score is 0.889"	
		/codon_start=1 /trans_table=11 /product="hypothetical protein Rv3497c"	
		/protein_id="CAA1773_4..1."	
		/db_xref="GI: 2924434"	
		/note="SPREML: 053542"	
		/translation="MURKPSKHEDPLRTGIGFLVLCVVLIAFGYSSCLPFWQGKTYDAYFTDAGGITPNSVYSGVKGAVASVLAGNAAKTVTSYDLSIVYQGDOSLAALNVFTQALHDDATPOVRLGAVDLTSRLRNDRDEAGLQLLAHAKSYTIVSLEAEQVINKLVEDQNLPAFDARAALSALESIDDDYAAQISCFVADRKERCPALSRLNLVIANLNERRDXTTEALKRKPITYATLGEVYVGSGPENVNVYSVLPGPLYATVFDLVFQPSKLPPDSLADYLRFQFQERNVILIRKPSP"	
		/complement(5272..6324)	
		/gene="Rv3498c"	
		/complement(5272..6324)	
		/note="Rv3498c, (MTV023_05c), len: 350. Unknown but similar to Mycobacterium tuberculosis proteins MTCT28_10 (346 aa), MTCT19H5_32 (346 aa) and MTV051_05. Hydrophobic region at N-terminus. FASTA scores: Z970501 MTCT28_10 (346 aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in 340 aa overlap; and Z97182 MTCT19H5_30 (508 aa) opt: 821 z-score: 697.0 E(): 2.6e-31; 35.1% identity in 453 aa overlap. TParse score is 0.891"	
		/codon_start=1 /trans_table=11 /product="lprN"	
		/protein_id="CAA1773_2."	
		/db_xref="GI: 2924432"	
		/note="SPREML: 053540"	
		/translation="MNTIWLRRIILPASSALLAGCGFGGLNSLPLPGTAGHGEAYSVTTEMADATLPONSPMVMDDTYTGSTAGIVAVORPDSFYAVKLDKNNVLPANAVAKVSQTSLLGSILVLLAPTRDRPCTLVDSRRIETRPTTEEVFSALGVVNNKGVNGALEBTTIDETHOAVAGRSQQFVNLYVPRDIAETLAGRNHRDHIDIDALDGLNRVSEAI LARDKDNIGRALDTLPDAVRLVNQNDRDHITDAAFKLRTAVTSHVLAETKVDFGEDL	

JOURNAL	kinase gene	VITRATAPNIGHTYSLDGVNIAKPTIVTSITYTGNCOCYRSATIRSYLDHPGHTQSCVCGSVERMYMAGAIDMLIYDDKDVPLQVAGENSTIPAFPKLYTSPMALLMFNGTVLMSRAFASAFKPTSTLWASTGGLLALILTVIVANLGGYTINNDYLINNE"
MEDLINE	90362067 (bases 1 to 5880)	polyA_signal 4777 .4782
REFERENCE	Nicolson,L.	/note="put. polyadenylation signal (gH); putative"
AUTHORS	Unpublished (1993)	repeat_region 4893 .5315
COMMENT	Submitted (10-SEP-1990) to DDBJ by:	/note="direct repeat region"
	Lesley Nicolson	rep_origin 5395 .5403
	Dept. Veterinary Pathology, University of Glasgow Vet School	/note="put. replication origin; putative"
	Beardsden Roda, Glasgow G61 1OH, Scotland UK.	rep_origin 5440 .5448
	Phone: 041-339-8855	/note="put. replication origin; putative"
	Fax: 041-330-5733	complement(5755 .>5880)
FEATURES	Location/Qualifiers	/note="putative; ORF4 (UL21 homologue)"
	1. .5880 /db_xref="GI:211824"	/codon_start=1
	1. .5880 /organelle="Equine herpesvirus 4"	/protein_id="BA03377.1"
	1. .5880 /db_xref="Raxon:10331"	/translation="NWSLTATFDNLDSRKGAYIYLYLSERLKRGGVPHVNRA"
source		BASE COUNT 1459 a 1377 c 1499 g 1545 t
ORIGIN		
CDS		Query Match 7.0%; Score 41.6%; Pred. No. 12; Length 5880; Best Local Similarity 4.1%; Mismatches 219; Indels 0; Gaps 0;
	1/codon_start=1	Qy 145 GCGACCCGAAACCGGGCTGGTCTCGGCACATGACGCCGAGTAGCAGTCATACC 204
	1/protein_id="GI:221821"	Db 4893 GCGGCCGAGGGCTGGCGGGAGGGCTGGCGGGAGGGCTGGCGGGCG 4952
	1/db_xref="GI:221821"	Qy 205 AATACGACTTCTGCCGCTGCGCATCCCCCTCATCACCCCTCATCACCCCTGTTCAGGCCAGCAAG 264
	1/translation="MRKQRQLTARSRLRAGTRCHARYNMAMQDLASAKNGYIGARLAPLSELVPAATLKTAMGVSAFEYNVQRPDCDTCVQHGSDAKGCLLIEIKTCRFSKNNNTASNLQKGMMRQHDSCRLLAKTLPGSGSEELTAPVLFVNQGMVRJVTLSQVVSINAALUSCTSLRAEISPPSERSSPRRCVTRRNUSKAFFAKITGSIQPIIQQARAAVASLFSATAQANTNAQAVQPATISLANPLAWASLFAPK" 861 . 865	Db 4953 GCGAGGCCCTGGCGGGGGCGAGGGCTCGGGCGGGGGGAGGGCTGGCGGGCG 5012
	1/note="put. TATA box (TK); putative"	Qy 265 GAGGCCGCCCCTGGCGGGATCGCTGAGCGGACGTGGGAGCAAGATCATT 324
	1/miss_feature 887 . 2007 /note="pot. transcription initiation site; putative"	Db 5013 GCGGCGAGGGCTGGCGGGGGCTGCGGGGGGAGGGCTGCTGGCGGGCTGTGCG 5072
	1/CDS /partial	Qy 325 GCTAACATTAGTGAATCAGCTGGGATTAATGTCACCCAAAAGGAGCTGTGACTGGA 384
	1/note="ORF2; (TK)" /codon_start=1	Db 5073 GCGGCCGAGGGCTGGCGGGGGCTGGCGGGGGAGGGCTG 5132
	1/protein_id="BA03378.1" /db_xref="GI:221822"	Qy 385 GCGCCGCCAGCAGCAAGTACGGGGTACGCTGGGACTACCCCGGAGCTGGCCACA 444
	1/translation="MAACPVGAEAPSASGPTRQVTVRIYLQVYGGKSTGRVMASAASCGSPTLYFSPPMVWTRLPEFTDVSIGYDQTNQRCQNIADDALITAHYOSRETPYLVLLDHTCULEFGGSIQLRQTOQDITVYDHRPASTVCEPANYILGYSYCALMANVATPREPOQGNIVVTTLVAEEHLRLTRARICBQDITLIALTRNYYFMLVNTCHFVEURSGWDRGEFLPTSCGAYKHATOMDAFOEVRSPEGLDTLFALFKIOELDDRGCHLVEIAWADALMILKRNLNVESADSLSGTPROCAVVESLPLMSLSDFDSSASALERARFAENAEGV" 2049 . 2054	Db 5133 GCGGCCGCGGGGGCTGGCGGGGGCTGCGGGCGGGCGAGGGCTG 5192
	1/note="put. polyadenylation signal (TK); putative"	Qy 445 AGTCATACGGCACATGGGACCTGTGAGGCCGACTGTGAGCTGAGTAATAGTGGT 504
	1/TATA_signal 2197 . 2202 /note="putative"	Db 5193 GCTGGCGGGGGCTGGCGGGGGCTGCTGGCGGGGGCTGCTGGCTGTGCG 5252
	1/CDS 2225 . 4792 /partial	Qy 505 GCCAGGCCGGCTCTGCCACAGGCCGAGGGCTG 536
	1/note="put. polyadenylation signal (TK); putative"	Db 5253 GCGAGGCCGAGGGCTGCGGGCGGGCTG 5284
	1/TATA_signal 2197 . 2202 /note="putative"	RESULT 9 RNU87960/c
	1/CDS 2225 . 4792 /partial	LOCUS RNU87960 2479 bp DNA ROD 04-MAR-1997
	1/note="ORF3; (gH)" /codon_start=1	DEFINITION Rattus norvegicus leukocyte common antigen receptor (LAR) gene, trans-spliced alternative untranslated exon.
	1/protein_id="BA03379.1" /db_xref="GI:221822"	ACCESSION U87960 1 GI:1857717
	1/translation="MSQPYKRIALYAAITIVSAIPWWTIPVSTSPQQOTKLHYVGNTRWNNTENVTYDRTMPEVNNNNSSTTTEVAIERNFETVNTLGASFWILKSNPPKHOQCIANPEPGDPRGCPVNSTVLFFNDLNFPLMTKNLFEVLDNYITGWTFERSKVATGPNPGVVLSPERTSDVNTLDEHTDPKHOHSLIDEHTVLDQFNTKLTLYLISPFAAWPTTAFAHAGITVNGCDTQIAYGNGFMQIISVNNPLEMFAPNDVRAVNLNPRLPPLRRLEPQPGYAGPIKYVLSUDGNYLGHOMSKIREVAYPEESLDYRHLSLNLDTLMAELSSLGSSYLYRIIIRLAVATESLAEVIRUDYMLQEAIDVDDNIRLPLIYPLVWKYARGTASSYSSDNDQFEEQAQKEIVADYIENELRKPMYBRSLSLXKSVAYSRKPLNAYSFANLITYKEAKDRITWNSTMREVLFFAVGAAGSHVLTGDPLGILGHAKHDSEMFSLNRNLLCTAMCFASHAVAGVKLEVNAGLIGGGVQFSLVEYFSPCMASRDLAEEVPLSLSVPPRTDINTGLEDDTIHSIGRSANGILNSRILAYNFDAVRVFTPELASCTKLPLVVLPLASNRY	KEYWORDS NORWAY rat, Rattus norvegicus
	1/REFERENCE 1 (bases 1 to 2479)	ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rat.
	1/AUTHORS Zhang C., Zhang,J.S., Martignetti,J.A., Massaa,S.M. and Longo,F.M.	REFERENCE 1 (bases 1 to 2479)
	1/JOURNAL Unpublished	AUTHORS Zhang C., Zhang,J.S., Martignetti,J.A., Massaa,S.M. and Longo,F.M.
	1/REFERENCE 2 (bases 1 to 2479)	REFERENCE 2 (bases 1 to 2479)
	1/AUTHORS Direct Submission	TITLE Direct Submission

JOURNAL	Submitted (30-JAN-1997) Neurology, Veterans Administration Medical Center at UC San Francisco, Neurology (127) 4150 Clement St., San Francisco, CA 94121, USA	
FEATURES	Location/Qualifiers	
source	1. .2479 /organism="Rattus norvegicus" /db_xref="Taxon:10116" <1..1649	
intron	/gene="LAR"	
gene	1..2479 /gene="LAR"	
repeat_region	1647..2456 / rpt_type=c tandem / rpt_unit=CAG 1650..2456 /gene="LAR"	
exon	/note="trans-spliced alternative untranslated exon" /product="Leukocyte common antigen receptor", 2457..2479 /gene="LAR"	
intron	832 a /base_count=504 c 584 g 559 t	

	Query	Match	Score	Length
	Best Local Matches	Similarity	41.2;	2479;
	Conservative	Pred. No.	DB 12;	
	73;	Mismatches	16;	
		Indels	0;	Gaps
b	173	CGGGCACTGACGGGGAGTAGCAGTATACCAATTACGACTTCTGCCTGCATCCC	6.9%	232
b	2186	CCTGCACTGTGCTGTACCTGTCTGCACTGTCTGCTGCTGCTGAGCTCGCTGCC	57.9%	2127
y	233	CTGATCACCCCTGTTACGGCCAGCGAACGGCCCTCTGCCCGCGCGAT	41.2;	292
b	2126	CTGAGCTGCCGTGGSCCTGGCTGCCCTGGCTGTCGGCTGGCCAGGCTGCC	41.2;	2067
y	293	CGGTG	298	
y	2056	CGCGT	2061	

REFERENCES
 1 (bases 1 to 477)
 Koob, M.D., Benzow, K.A., Bird, T.D., Moseley, M.L. and Ranum, L.P.W.
 AUTHORS
 TITLE
 Rapid cloning of expanded trinucleotide repeat sequences
 JOURNAL
Nature Genet. (1997) In Press
 2 (bases 1 to 477)
 Koob, M.D., Benzow, K.A., Moseley, M.L. and Ranum, L.P.W.
 AUTHORS
 TITLE
 Direct Submission
 JOURNAL
 Submitted (19-AUG-1997) Neurology, University of MN, Box 295 UMHC,
 416 Delaware St. SE, Minneapolis, MN 55455, USA
 FEATURES
 Location/Qualifiers
 1. 477
 SOURCE

```

<?xml version="1.0" encoding="UTF-8"?>
<organism>Homo sapiens</organism>
<db_xref>taxon:9606</db_xref>
<chromosome>3</chromosome>
<map>p</map>
<note>isolated from a spinocerebellar ataxia 7 (SCA7) patient</note>
<patient>115..47</patient>
<note>similar to NotI jumping clone sequence with GenBank Accession Number X95831</note>
<repeat_region misc_feature="repeat region">
<repeat_id>127</repeat_id>
<repeat_start>295</repeat_start>
<repeat_end>295</repeat_end>
<repeat_length>1</repeat_length>
<repeat_gc_content>0.47</repeat_gc_content>
<repeat_nucleotide>C</repeat_nucleotide>
</repeat_region>

```

```
/note="region of cag expansion"  
/rpt_unit=cag  
99 a 164 c 172 g 42 t  
BASE COUNT  
ORIGIN
```

```

/tissue_type="brain"
/clone_lib="cDNA library (custom made by Strategene)"
<1..1863
/codon_start=1
/product="leucocyte common antigen-related protein"
/protein_id="CAA8537.1"
/db_xref="SPTRREMBL:Q63296"
/translation="STIGKDSLLAHSSDPVEMRRLNQTPGSSAPSCPNISSMRDHPP
ITPDLDNIEERIKAQDGLKFSEDEYESIDPGQOFTWENNSEYVKPKRNAYAVIAYDH
SRVLITISDGPGSDYIANYIDGYRKONAYIATQGPFLETMIDFWRMWEORTATVV
MNTTRLEFKSRVYKCDQXPKPARGEFETYGLQVLTUDTVEYLTYTMATEFLAKGSSESKRE
LRQFQENAWPDIGDVPEFTPILAFLLRKYACNPIDQGPVPHCSAGVGIGFCIVIDA
MLERMHEKTVTDLYGHYTCMRSQNYMQTFIHFVHLLAECMAMCQKRNLYVNIMPYLN
AHIQKIGQVOPPCESVTAMLEFFKLANSKAHTSRSFVI SANLPCNKFKNRVLNIMPYLN
AVCLQPQFVGESDYINASLDYRQKAYIATQGPPLAESTEDFWRLWHNSTIIVM
LTKLRENGRKCHQYWAERSARYQFVDPMAYINMPQYIIRFEKVTDRTGQFSTI
RQEFUDWMPRECYPKTICQYVWVQYQFEGODQGQFQDGTGQFSTI
IVLERYRYEGVYDMFQVYKTLQRQPMVQTEDOYQLCYRAALEYLGSFDHYAT
80..112
misc_feature
  /note="insertion sequence LAS Ea"
  3' UTR
  2104..3941
  /note="alternate spliced 3' UTR with CAG repeats"
  BASE COUNT 1001 a 1069 c 1061 g 810 t
  ORIGIN

```

AUTHORS Parkhill, J.

TITLE Direct Submission

JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unit de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk

COMMENT On Jun 20, 1998 this sequence version replaced gi:2661623.

Notes:

- details of *M. tuberculosis* sequencing at the sanger Centre are available on the World Wide Web.
- ([URL](http://www.sanger.ac.uk/Projects/M_tuberculosis/), http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
- Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
- CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES	Location/Qualifiers
source	1. .63350 "Mycobacterium tuberculosis" /strain="H37Rv" /db_xref="taxon:1773"
misc_feature	<1. .68107 /note="fragment designated v004 . Does not represent a physical clone." complement:144 . 401)
gene	/gene="Rv3346c"
repeat_unit	144 . 350 /note="direct repeat 1" complement:144 . 401)
CDS	/gene="Rv3346C" /note="Rv3346C, (MTV004_02c), len: 85, unknown, similar to M. leprae hypothetical protein TR.E332051 (BMBL_298271) MLCB1779_16C (91 aa) fasta scores: opt: 349 z-score: 701.3 E(1): 6.6e-32, 67.4% identity in 95 aa overlap. Also highly similar to upstream ORF Rv3355c (MTV004_12c) (85.6% identity in 97 aa overlap)" /codon_start=1 /transl_table=11 /product="hypothetical protein Rv3346C" /protein_id="CAA15731_1" /db_xref="GI: 2661625" /db_xref="SPTRNBL:OS030377" /translation="MTVRARVRRTVGAQWPFLAGYNEWRRGALLIGTGVAALRLVLSSEERAGLVYRSKGIDFVTIVTVAAMMVLASTIDPLGTG"
repeat_unit	348 . 456 /note="direct repeat 2" complement:657 . 24048)
repeat_unit	454 . 551 /note="direct repeat 3" complement:657 . 24048)
gene	/gene="PPE"
CDS	/complement(657 . 10130) /gene="PPE" /note="Rv3347C, (MTV004_03c), member of the M. tuberculosis PPE family of Gly - Ala - Asn-rich proteins, len: 3157 aa; similar to many eg. TR:00732 (BMBL_296800) MTCTY63_10c (963 aa) opt: 1181 z-score: 2191.5 E(1): 0, 47.5% identity in 995 aa overlap. Also similar to upstream ORF MTV304_5, MTY13E10_16, MTCTY28_16, MTCTY63_9, MTY13E10_17, MTCTY180_1 etc." /codon_start=1 /transl_table=11 /product="PPE" /protein_id="CAA15732_1" /db_xref="GI: 2661626" /protein_id="CAA15732_1"

/note="Rv3348, (MTY004_04), len: 163 aa; unknown,
identical to M. tuberculosis hypothetical protein
TR:296234 (EMBL:Z83864) MTC01AG_25 (163 aa). Partially
similar to several insertion elements, e.g.
P19334|Y111_STRL INSERTION ELEMENTIS116, (399 aa, 28.5%
identity in 158 aa overlap)."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3348"
/protein_id="CAA1513_3.1"
/db_xref="GI:2661627"
/db_xref="SPOTMBL:P96234"
/translation="MTAENPGRSRTLVNGIDAATACHHIAIRDVGARSIRPSVEPT
LAGIRLTDLKSLGDDIDATEVPSMTWLPITAVENGDMMAGASHCARURGAIV
GKSFSDVDAEVLRASEVFDLTPLTFAQLLRSVIRRASVIDANSRRILMS
LAR"
gene
/gene="Rv3349c"
complement("11239 . .11979")
CDS
/gene="Rv3349c"
/note="Rv3349c, (MTY004_05C), len: 246 probable
transposase Pseudogene fragment, similar to part of
TR:050911 (EMBL:U10634) IS204 PUTATIVE TRANSPOSE from
NOCARDIA ASTEROIDES (377 aa), fasta scores: opt: 288
-score: 322.6 E(): 8.3e-11, 48.5% identity in 97 aa
overlap"
/codon_start=1
Query Match Score 6.8%; Best Local Similarity 53.1%; Length 69350;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
CvY 142 GTGGGCGACGCGAAACGGGGCTCTGGGCACACTGGGGCAACTGGAGTCAT 201
Db 37024 GCGGCCACGGGGTGA CGGGGGCATGCCGCACGGTGA CGGTACTGGCCG 37083
Qy 202 ACCAATACGAGCTTCTCCGGCTGTCTGATCCCTGTGATCACCCCTGTGTTAGGCCAGGCC 261
Db 37084 GCCAGGACGCCAACCGGGATGGGGATGGGGTGGGGCACCC 37143
Qy 262 AAGGAGCGGGCTCTGGGGGGATGGGTGGGGTGAAGCCG 303
Db 37144 ATGGCACCCCCGGAGGGCAACGZAGGGCACCGCACCG 37185
RESULT 14
AP000556/c AP000556 149618 bp DNA PRI 28-OCT-1999
DEFINITION Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region,
clone:KB1172b5.
ACCESSION AP000556 2 GI:6139049
VERSION 1
KEYWORDS Homo sapiens DNA, clone:KB1172D5.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Shimizu.N.
TITLE Human DNA sequence from clone KB1172D5 on chromosome 22q11.2
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 1 (bases 1 to 14918)
AUTHORS Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan
(E-mail: nshimizu@dnb.med.keio.ac.jp, Tel:81-3-3351-2370,
Fax:81-3-3351-2370)
COMMENT On Oct 29, 1999 this sequence replaced g1:6006352.
This is a complete sequence of the insert of clone KB1172D5 with 67422 bp
proximal adjacent clone is KB118305 (Acc.#AP000552) with 67422 bp
overlapping. The distal adjacent clone is KB1323b2 (Acc.#AP000557)
FEATURES with 69793-bp overlapping.
source Sequence updated (26-Oct-1999).
Location/Qualifiers
1 . 149618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="KB1172D5"
/map="22q11.2"
repeat_region complement("11239 . .11979")
repeat_region /rpt_family="L1M4C"
/evidence="not_experimental complement(10_176)"
/rpt_family="L1M4C"
repeat_region /rpt_family="L1M4C"
/evidence="not_experimental complement(1717 . .2021)"
repeat_region /rpt_family="AluSx"
/evidence="not_experimental complement(2022 . .2297)"
repeat_region /rpt_family="L1M4C"
/evidence="not_experimental complement(229 . .2368)"
repeat_region /rpt_family="L1MB4"
/evidence="not_experimental complement(2766 . .3043)"
repeat_region /rpt_family="AluSg"
/evidence="not_experimental complement(3044 . .3066)"
repeat_region /rpt_family="TAAA)n"
/evidence="not_experimental complement(4122 . .4305)"
repeat_region /rpt_family="AluSg(x)"
/evidence="not_experimental complement(4122 . .4305)"
repeat_region /rpt_family="AluJb"
/evidence="not_experimental complement(4949 . .5151)"
repeat_region /rpt_family="AluJ"
/evidence="not_experimental complement(5619 . .5748)"
repeat_region /rpt_family="MIR"
/evidence="not_experimental complement(8880 . .9173)"
repeat_region /rpt_family="AluY"
/evidence="not_experimental complement(10813 . .10860)"
repeat_region /rpt_family="GC-rich"
/evidence="not_experimental complement(10822 . .10892)"
repeat_region /rpt_family="CGG)n"
/evidence="not_experimental complement(11091 . .11118)"
repeat_region /rpt_family="GC-rich"
/evidence="not_experimental complement(11968 . .12299)"
repeat_region /rpt_family="AluJ"
/evidence="not_experimental complement(13431 . .13529)"
repeat_region /rpt_family="MIR"
/evidence="not_experimental complement(15901 . .16312)"
repeat_region /rpt_family="MER31B"
/evidence="not_experimental complement(17091 . .17265)"
repeat_region /rpt_family="L12"
/evidence="not_experimental complement(1729 . .17272)"
repeat_region /rpt_family="MIR"
/evidence="not_experimental complement(17591 . .17671)"

```

/rept_family="G-rich"
/evidence="not_experimental
complement(1767. .17831)
/rept_family="AluSx"
/evidence="not_experimental
17832 .17880
/rept_family="MER45C"
/evidence="not_experimental
21993 .22032
/rept_family="Aluy"
/evidence="not_experimental
22304 .22373
/rept_family="L1MD3"
/evidence="not_experimental
22888 .23013
/rept_family="AluYb8"
/evidence="not_experimental
23014 .23323
/rept_family="Aluy"
/evidence="not_experimental
23324 .23496
/rept_family="AluYb8"
/evidence="not_experimental
complement(24098 .24406)
/rept_family="AluUb"
/evidence="not_experimental
complement(24447 .24558)
/rept_family="MIR"
/evidence="not_experimental
24631 .24721
/rept_family="CT-rich"
/evidence="not_experimental
24952 .25245
/rept_family="AluSx"
/evidence="not_experimental
26011 .26140
/rept_family="L2"
/evidence="not_experimental
26310 .26339
/rept_family="(TTTG)n"
/evidence="not_experimental
complement(26340 .26623)
/rept_family="AluSx"
/evidence="not_experimental
26731 .26813
/rept_family="L2"
/evidence="not_experimental
complement(26798 .26898)
/rept_family="L2"
/evidence="not_experimental
26899 .27029
/rept_family="FLAM_C"
/evidence="not_experimental
complement(27037 .27337)
/rept_family="Aluy"
/evidence="not_experimental
complement(27338 .27380)
/rept_family="L2"
/evidence="not_experimental
27868 .28160
/rept_family="Aluy"
/evidence="not_experimental
complement(28160 .28255)
/rept_family="(TTTAA)n"
/evidence="not_experimental
complement(28298 .28602)
/rept_family="AluUb"
/evidence="not_experimental
complement(28612 .28912)
/rept_family="AluSx"
/evidence="not_experimental
complement(28927 .29192)
/rept_family="AluSg"

```

repeat_region /evidence="not_experimental
repeat_region complement(29229 .29542)
repeat_region /rpt_family="AluSx"
repeat_region /evidence="not_experimental
repeat_region complement(29544 .29741)
repeat_region /rpt_family="7SLRNA"
repeat_region /evidence="not_experimental
29808 .30107
repeat_region /rpt_family="AluSx"
repeat_region /evidence="not_experimental
repeat_region complement(30206 .30372)
repeat_region /rpt_family="L2"
repeat_region /evidence="not_experimental
repeat_region complement(30521 .30840)
repeat_region /rpt_family="Aluy"
repeat_region /evidence="not_experimental
repeat_region complement(30841 .30969)
repeat_region /rpt_family="AluUb"
repeat_region /evidence="not_experimental
repeat_region complement(31012 .31228)
repeat_region /rpt_family="AluJO"
repeat_region /evidence="not_experimental
repeat_region complement(31307 .31387)
repeat_region /rpt_family="MIR"
repeat_region /evidence="not_experimental
31540 .31819
repeat_region /rpt_family="AluSx"
repeat_region /evidence="not_experimental

Query Match 6.8%; Score 40.4; DB 10; Length 149618;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 109; Conservative 0; Mismatches 96; Indels 1; Gaps 1;

Qy	141	AGTGGCCAGGGGAACGGGGTGTGTCCTCGGCCACTGACGCCGAGTAGCAGTCA 200
Db	35731	AGCAGGGCCGGCTGCAGCTGCGGCTCTCGCCACSTCGCCGCGGGCCGCC 35672
Qy	201	TACCAATACGACTTCGCTGCATCCCTGCATACCCCCCTGTCTAGGCCAGC 260
Db	35671	TGGTGGCCAGGCTCCGGACTCGGCCGGCTGGCTCAGGCCAGC 35612
Qy	261	CAAGGCAGCCGGCCMTCCTGCCGGCGGATC-GGCTGAGCCGACGTGGGGAGAAGA 319
Db	35511	ATCTGCCGTGCGACCTCTCGAGGTGCGCTGCGCCCTGGCCAAGGCCA 35552
Qy	320	TCAATTGCTTAATCTAGTGAATAGCTG 345
Db	35551	CCGGTTGCCCTGTGTGCGATCAAATG 35526

RESULT 15

AP000557	150036 bp DNA	PRI 02-NOV-1999
LOCUS	Homo sapiens genomic DNA, chromosome 22q11.2, BCRU2 region,	DEFINITION
DEFINITION	Clone:KB1323B2.	DEFINITION
ACCESSION	AP000557	ORGANISM
VERSION	AP000557.2	KEYWORDS
KEYWORDS	Homo sapiens DNA, clone:KB1323B2.	KEYWORDS
SOURCE	Homo sapiens	REFERENCE
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	REFERENCE
ORGANISM	Eutheria; Primates; Catarrhini; Hominidae; Homo.	REFERENCE
VERSION	1 (bases 1 to 150036)	REFERENCE
VERSION	AP000557.2 GI:6174873	REFERENCE
VERSION	Shimizu,N.	AUTHORS
VERSION	Homo sapiens genomic DNA, chromosome 22q11.2, BCRU2 region, clone	AUTHORS
VERSION	KB1323B2	AUTHORS
VERSION	Published Only in DataBase (1999) In press	JOURNAL
VERSION	2 (bases 1 to 150036)	JOURNAL
VERSION	Shimizu,N.	AUTHORS
VERSION	Direct Submission	TITLE
VERSION	Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular	TITLE
VERSION	Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan	TITLE

(E-mail:nshimizu@ddmb.med.keio.ac.jp, Tel:81-3-3351-2370;
 Fax:81-3-351-2370)
 On Nov 2, 1999 this sequence version replaced gi|6006353.
 This is a complete sequence of the insert of KB123B2 clone. The
 proximal adjacent clone is KB117D5 (Acc. #AB00056) with 69794 bp
 overlapping. The distal adjacent clone is KB1802C5 (Acc. #AB000558)
 with 21623 bp overlapping.
 Sequence updated (29-Oct-1999).

FEATURES

source 1. 150036 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="KB1323B2"
 /map="22q11.2"
 /rpt_family="AluSx"
 /evidence="not_experimental complement(519..594)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(554..565)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(566..594)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(595..716)"
 repeat_region /rpt_family="FLAM_C"
 /evidence="not_experimental complement(813..1106)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(1918..2204)"
 repeat_region /rpt_family="AluJb"
 /evidence="not_experimental complement(2613..2915)"
 repeat_region /rpt_family="Alusp"
 /evidence="not_experimental complement(3253..3333)"
 repeat_region /rpt_family="LIME3A"
 /evidence="not_experimental complement(3396..3501)"
 repeat_region /rpt_family="ME5A"
 /evidence="not_experimental complement(3779..4072)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(4132..4429)"
 repeat_region /rpt_family="AluSx"
 /evidence="not_experimental complement(5553..5655)"
 repeat_region /rpt_family="ME5A"
 /evidence="not_experimental complement(5930..6220)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(7330..7497)"
 repeat_region /rpt_family="MIR"
 /evidence="not_experimental complement(9725..9855)"
 repeat_region /rpt_family="LIME3A"
 /evidence="not_experimental complement(9901..10191)"
 repeat_region /rpt_family="AluY8"
 /evidence="not_experimental complement(17988..18048)"
 repeat_region /rpt_family="(Tg)n"
 /evidence="not_experimental complement(22321..22478)"
 repeat_region /rpt_family="HALL"
 /evidence="not_experimental complement(22670..22733)"
 repeat_region /rpt_family="(TA)n"
 /evidence="not_experimental complement(22743..22997)"
 repeat_region /rpt_family="FLAM_C"

/rpt_family="AluSx"
 /evidence="not_experimental complement(22938..23084)"
 repeat_region /rpt_family="MBR58A"
 /evidence="not_experimental complement(23091..23406)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(23402..23550)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(23612..23757)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(23758..24046)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(24047..24058)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(24059..24357)"
 repeat_region /rpt_family="AluSx"
 /evidence="not_experimental complement(24358..24506)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(24507..24815)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(24816..24846)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(24847..25133)"
 repeat_region /rpt_family="AluJb"
 /evidence="not_experimental complement(26278..26587)"
 repeat_region /rpt_family="AluSx"
 /evidence="not_experimental complement(26985..27136)"
 repeat_region /rpt_family="AluJb"
 /evidence="not_experimental complement(26802..26972)"
 repeat_region /rpt_family="L2"
 /evidence="not_experimental complement(27136..27322)"
 repeat_region /rpt_family="AluJb"
 /evidence="not_experimental complement(27951..28039)"
 repeat_region /rpt_family="L2"
 /evidence="not_experimental complement(28041..28351)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(28352..28383)"
 repeat_region /rpt_family="(RTG)n"
 /evidence="not_experimental complement(28384..28664)"
 repeat_region /rpt_family="AluJb"
 /evidence="not_experimental complement(28669..28964)"
 repeat_region /rpt_family="AluJ"
 /evidence="not_experimental complement(28968..29280)"
 repeat_region /rpt_family="AluJb"
 /evidence="not_experimental complement(29359..29484)"
 repeat_region /rpt_family="AluSg"
 /evidence="not_experimental complement(29487..29566)"
 repeat_region /rpt_family="AluY"
 /evidence="not_experimental complement(29606..29747)"
 repeat_region /rpt_family="FLAM_C"

```

repeat_region          /evidence=not_experimental
2950 . 30039
/rept_family="AluSg"
/evidence=not_experimental
repeat_region          30042 . 30356
/rept_family="AluJo"
/evidence=not_experimental
complement(30945 . 31057)
/rept_family="MUTIC"
/evidence=not_experimental
complement(31068 . 31371)
/rept_family="AluYb8"
/evidence=not_experimental
complement(31372 . 31489)
/rept_family="MUTIC"
/evidence=not_experimental
complement(31508 . 31792)
/rept_family="AluYb"
/evidence=not_experimental
31892 . 32012
/rept_family="AT_Rich"
/evidence=not_experimental
complement(32013 . 32138)
/rept_family="FLAM_C"
/evidence=not_experimental
32285 . 32279
/rept_family="L1M4"
/evidence=not_experimental
3228 . 32967
/rept_family="AluJb"
/evidence=not_experimental
complement(32986 . 33124)
/rept_family="MUTIC"

Query Match           6 . 8%   Score 40 . 4;  DB 10;  Length 150036;
Best Local Similarity 52 . 9%;  Pred: No 17;  Gaps 1;
Matches 109;  Conservative 0;  Mismatches 96;  Indels 1;
Db 119367 AGCGAGGCCGAAACGGGGTGTGTCCTGGGGCACTGAGCGGGGAGTAGCAGTC 200
Qy  141 ATGGCGAGCCAGCTGGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 119426
Db 119427 TGGTGCCAGCTGGACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 260
Qy  201 TACCAAATAGACTCTGGCTGGCTGGCATCCCTGTCATCACCCCTGTTCAAGGCCAGC 119486
Db 119487 ATCTGGCTGCACTCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 319
Qy  320 TCAATTGCTTAATCTAGTGAATCAAGCTG 345
Db 119547 CGTGTGCCCTGCTGAAGCATCAACTG 119572

```

Gencore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 22:08:30 ; Search time 66.13 Seconds
(without alignments)
1075.390 Million cell updates/sec

Title: US-08-988-242-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 CAGGTACAGCTAACGGCTT.....AACGAGCTCGGGATA 594

Scoring table: IDENTITY_NUC
Gapp 10.0 , Gapext 1.0

Searched: 21494 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/pctodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/pctodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/pctodata/1/ina/5C_COMB.seq:*
- 4: /cgn2_6/pctodata/1/ina/5D_COMB.seq:*
- 5: /cgn2_6/pctodata/1/ina/6_COMB.seq:*
- 6: /cgn2_6/pctodata/1/ina/pct059_COMBO.seq:*
- 7: /cgn2_6/pctodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	5.94	100.0	3402	2	US-08-480-917-1	Sequence 1, Appli
2	39.2	6.6	2790	3	US-08-718-661-1	Sequence 1, Appli
3	38.2	6.4	2943	1	US-08-042-747A-7	Sequence 7, Appli
4	37.6	6.3	2415	4	US-08-474-379C-60	Sequence 6, Appli
5	36.6	6.2	913	1	US-08-217-322-3	Sequence 3, Appli
6	36.6	6.2	913	1	US-07-885-970A-3	Sequence 3, Appli
7	36.6	6.2	913	1	US-08-288-687A-3	Sequence 3, Appli
8	36.6	6.2	913	1	US-08-510-797-2	Sequence 2, Appli
9	36.6	6.2	913	1	US-08-288-829-3	Sequence 3, Appli
10	36.6	6.2	913	4	US-08-717-335-2	Sequence 25, Appli
11	36.6	6.2	1984	1	US-07-855-970A-25	Sequence 25, Appli
12	36.6	6.2	1985	1	US-08-288-687A-25	Sequence 25, Appli
13	36.6	6.2	1985	1	US-08-288-829-25	Sequence 25, Appli
14	36	6.1	5452	4	US-09-110-114-1	Sequence 1, Appli
C 15	36	6.1	10596	1	US-07-894-811-15	Sequence 15, Appli
C 16	36	6.1	10596	1	US-07-855-971-15	Sequence 15, Appli
C 17	36	6.1	10596	1	US-08-037-782A-15	Sequence 15, Appli
C 18	36	6.1	10596	1	US-08-194-088B-15	Sequence 15, Appli
C 19	36	6.1	10596	1	US-08-194-088B-15	Sequence 15, Appli
C 20	36	6.1	10596	6	PCT-US93-0468-15	Sequence 15, Appli
C 21	35.2	5.9	3376	1	US-08-320-559-29	Sequence 29, Appli
C 22	35.2	5.9	3376	6	PCT-US94-04496-29	Sequence 29, Appli
C 23	35	5.9	168	2	US-08-469-802B-4	Sequence 4, Appli
C 24	35	5.9	168	3	US-08-267-803B-4	Sequence 4, Appli
C 25	35	5.9	171	3	US-08-469-802B-5	Sequence 5, Appli
C 26	35	5.9	171	3	US-08-267-802B-5	Sequence 5, Appli
C 27	35	5.9	195	2	US-08-469-802B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-480-917-1
; Sequence 1, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOES-BACCALA, Glaucia
; LESENEHAL, Mylene
; JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPANOSAMA CRUZI ANTIGEN, AND GENE
; ENCODING THE LATTER, THEIR APPLICATION TO THE DETECTION OF
; CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Olliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,917
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36400
; TELECOMMUNICATION INFORMATION:
; CLASSIFICATION: 435
; TELEPHONE: 703-36-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-480-917-1
Qy 1 CAGGTACAGCTAACGGCTTGCATCAATCGTAGCCGACGGTAGCTCGTCCTGGCT 60

Db 1232 CAGGTACACGGTAAACGGCTTTGTTCAATCGTAACAGCGAAGGTAGTCGCTCCCTGCCTGCT 1291
 Qy 61 GACATGTCATGGATTGAGAATCTCCGTCATTCAGGAAAGAACAGCAG 120
 Db 1292 GACATGTCATGGATTGAGAATCTCCGTCATTCAGGAAAGAACAGCAG 1351
 Qy 121 CAGGCCAAAACATCGTAGTCGACGGAAACCGGGTGTTCTCGGGCACT 180
 Db 1352 CAGGCCAAAACATCGTAGTCGACGGAAACGGGGTGTTCTCGGGCACT 1411
 Qy 181 GAGGGGGAGTAGCAGTCATACCAATACGACTCTGCCTGCATCCCTGCATCA 240
 Db 1412 GACCGGGAGTAGCAGTCATACCAATACGACTCTGCCTGCATCCCTGCATCA 1471
 Qy 241 CCCCTGTTTACGGCCAGCAAGCACCCGGCTCTGGCCGGCATGGCTGAG 300
 Db 1472 CCCCTGTTTACGGCCAGCAAGCACCCGGCTCTGGCCGGCATGGCTGAG 1531
 Qy 301 CGCACGTTGGACAAAGATAATTGCTAACTAGTGAATCAGCTGGGATTATGTACC 360
 Db 1532 CGCACGTTGGACAAAGATAATTGCTAACTAGTGAATCAGCTGGGATTATGTACC 1591
 Qy 361 CAAGGAGCGTCGTCAGGACTCTGGGAGGGCTACGGGGTGACGCTC 420
 Db 1592 CAAGGAGCGTCGTCAGGACTCTGGGAGGGCTACGGGGTGACGCTC 1651
 Qy 421 ACGACTACGCCCGAGGAAACAGTCATAGGCAATGGGACAATGGGACCTTGACGGCT 480
 Db 1652 ACGACTACGCCCGAGGAAACAGTCATAGGCAATGGGACAATGGGACCTTGACGGCT 1711
 Qy 481 GATTGGTGGAGCTTAATAGTGGCCAGGCCGCTCTGGCCACAGCCCAGGAA 540
 Db 1712 GATTGGTGGAGCTTAATAGTGGCCAGGCCGCTCTGGCCACAGCCCAGGAA 1771
 Qy 541 CAAACAGGAGAAAAGGCTCCGGCATGTGAAACGGGCTCGTGGGATA 594
 Db 1772 CAAACAGGAGAAAAGGCTCCGGCATGTGAAACGGGCTCGTGGGATA 1825

RESULT²
 ; Sequence 1, Application US/08718661
 ; GENERAL INFORMATION:
 ; APPLICANT: Nucleic acid molecules coding for mammalian
 ; TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation
 ; NUMBER OF SEQUENCES: 15
 ; CURRENT APPLICATION DATA:
 ; FILING DATE:
 ; US-08-718-661-1
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 ; APPLICATION NUMBER: US/08/718,661
 ; FILING DATE:
 ; US-08-718-661-1
 ; Sequence 1, Application US/08718661
 ; GENERAL INFORMATION:
 ; APPLICANT: Nucleic acid molecules coding for mammalian
 ; TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation
 ; NUMBER OF SEQUENCES: 15
 ; CURRENT APPLICATION DATA:
 ; FILING DATE:
 ; US-08-718-661-1
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 ; APPLICATION NUMBER: US/08/718,661
 ; FILING DATE:
 ; US-08-718-661-1
 ; CLASSIFICATION: 530
 ; FEATURE: sequence characteristics:
 ; NAME/KEY: CDS
 ; LENGTH: 2790 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 542 .. 2545

Query Match Score 38.2; DB 1; Length 2943;
 Best Local Similarity 47.7%; Pred. No. 0.11; Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
 Query 209 CGACTTCGCCCCCTGCTGATCCCTGATCAGGGCAGGCCAAGGCAG 268
 Db 255 CGCCCTGCTCCGGGACCTCTGGCTGGGGCTGGGAGCGGGCTGGCG 196
 Query Match Score 39.2; DB 3; Length 2790;
 Best Local Similarity 54.1%; Pred. No. 0.059; Matches 111; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

RESULT³
 ; Sequence 7, Application US/08042747A
 ; Patent No. 5487969
 ; GENERAL INFORMATION:
 ; APPLICANT: Eberle, Richard
 ; APPLICANT: Black, Darla
 ; APPLICANT: Scinariello, Franco
 ; APPLICANT: Hilliard, Julia K.
 ; TITLE OF INVENTION: Cloning and Amplification of Monkey B
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cox & Smith Incorporated
 ; STREET: 112 East Pecan Street, Suite 2000
 ; CITY: San Antonio
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 78205
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Halimond, W. Bradley
 ; REGISTRATION NUMBER: 35186
 ; REFERENCE/DOCKET NUMBER: S-0072-179
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 210-554-5500
 ; TELEFAX: 210-226-8395
 ; TELEX: 767609
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2943 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 87 .. 2744
 ; US-08-042-747A-7

Query Match Score 38.2; DB 1; Length 2943;
 Best Local Similarity 47.7%; Pred. No. 0.11; Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Query 269 CGACTTCGCCCCCTGCTGATCCCTGATCAGGGCAGGCCAAGGCAG 268
 Db 255 CGCCCTGCTCCGGGACCTCTGGCTGGGGCTGGGAGCGGGCTGGCG 196
 Query Match Score 39.2; DB 3; Length 2790;
 Best Local Similarity 54.1%; Pred. No. 0.059; Matches 111; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 258 AGCCAAGGGAGCCGCCCTCGCGGGGATGGCTGAGCCGC 304
Db 380 ACCTCCAGTTCTCACCTTCGCACTCCTCCACCGCTTCTCCAC 426

RESULT 6
US-07-885-970A-3

Sequence 3, Application US/07885970A

Patent No. 5495070

GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nicholas J. Seay, Quarles & Brady

STREET: P.O. Box 2113, First Wisconsin Plaza

CITY: Madison

STATE: Wisconsin

COUNTRY: USA

ZIP: 53701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/885, 970A

FILING DATE: 19920518

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/617, 239

FILING DATE: 21 NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/253, 243

FILING DATE: 04-OCT-1988

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27, 396

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 283-2478

TELEFAX: (608) 251-5139

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ANTI-SENSE: NO

ORIGINAL SOURCE: Gossypium hirsutum

STRAIN: Coker 312

DEVELOPMENTAL STAGE: 15 day old fiber cells

TISSUE TYPE: fiber cells

IMMEDIATE SOURCE:

LIBRARY: CKFB15A1

CLONE: H6

US-07-885-970A-3

RESULT 7
US-08-298-687A-3

Sequence 3, Application US/08298687A

Patent No. 552,078

GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nicholas J. Seay, Quarles & Brady

STREET: P.O. Box 2113, First Wisconsin Plaza

CITY: Madison

STATE: Wisconsin

COUNTRY: USA

ZIP: 53701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/298, 687A

FILING DATE: 04-OCT-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/617, 239

FILING DATE: 21 NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/253, 243

FILING DATE: 04-OCT-1988

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27, 396

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 283-2478

TELEFAX: (608) 251-5139

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ANTI-SENSE: NO

ORIGINAL SOURCE: Gossypium hirsutum

STRAIN: Coker 312

DEVELOPMENTAL STAGE: 15 day old fiber cells

TISSUE TYPE: fiber cells

IMMEDIATE SOURCE:

LIBRARY: CKFB15A1

CLONE: H6

US-08-298-687A-3

RESULT 198
US-08-298-687A-3

Sequence 3, Application US/08298687A

Patent No. 552,078

GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: PLANTS FOR ALTERED FIBER

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nicholas J. Seay, Quarles & Brady

STREET: P.O. Box 2113, First Wisconsin Plaza

CITY: Madison

STATE: Wisconsin

COUNTRY: USA

ZIP: 53701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/298, 687A

FILING DATE: 04-OCT-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/617, 239

FILING DATE: 21 NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/253, 243

FILING DATE: 04-OCT-1988

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27, 396

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 283-2478

TELEFAX: (608) 251-5139

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ANTI-SENSE: NO

ORIGINAL SOURCE: Gossypium hirsutum

STRAIN: Coker 312

DEVELOPMENTAL STAGE: 15 day old fiber cells

TISSUE TYPE: fiber cells

IMMEDIATE SOURCE:

LIBRARY: CKFB15A1

CLONE: H6

US-08-298-687A-3

RESULT 258
US-08-298-687A-3

Sequence 3, Application US/08298687A

Patent No. 552,078

GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: PLANTS FOR ALTERED FIBER

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nicholas J. Seay, Quarles & Brady

STREET: P.O. Box 2113, First Wisconsin Plaza

CITY: Madison

STATE: Wisconsin

COUNTRY: USA

ZIP: 53701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/298, 687A

FILING DATE: 04-OCT-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/617, 239

FILING DATE: 21 NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/253, 243

FILING DATE: 04-OCT-1988

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27, 396

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 283-2478

TELEFAX: (608) 251-5139

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ANTI-SENSE: NO

ORIGINAL SOURCE: Gossypium hirsutum

STRAIN: Coker 312

DEVELOPMENTAL STAGE: 15 day old fiber cells

TISSUE TYPE: fiber cells

IMMEDIATE SOURCE:

LIBRARY: CKFB15A1

CLONE: H6

US-08-298-687A-3

RESULT 320
US-08-298-687A-3

Sequence 3, Application US/08298687A

Patent No. 552,078

GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: PLANTS FOR ALTERED FIBER

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nicholas J. Seay, Quarles & Brady

STREET: P.O. Box 2113, First Wisconsin Plaza

CITY: Madison

STATE: Wisconsin

COUNTRY: USA

ZIP: 53701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/298, 687A

FILING DATE: 04-OCT-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/617, 239

FILING DATE: 21 NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/253, 243

FILING DATE: 04-OCT-1988

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27, 396

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 283-2478

TELEFAX: (608) 251-5139

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ANTI-SENSE: NO

ORIGINAL SOURCE: Gossypium hirsutum

STRAIN: Coker 312

DEVELOPMENTAL STAGE: 15 day old fiber cells

TISSUE TYPE: fiber cells

IMMEDIATE SOURCE:

LIBRARY: CKFB15A1

CLONE: H6

US-08-298-687A-3

RESULT 320
US-08-298-687A-3

Sequence 3, Application US/08298687A

Patent No. 552,078

GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: PLANTS FOR ALTERED FIBER

US-08-530-797-2
; Sequence 2, Application US/08530797
; Patent No. 559718
GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; Umbreck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18
; ADDRESSEE: Charles and Brady
; STREET: P.O. BOX 2113
; STATE: FIRST WISCONSIN PLAZA
; CITY: MADISON
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 800kb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,797
; FILING DATE: 20-SEP-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/617,239
; FILING DATE: 21-NOV-90
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-88
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
; US-08-530-797-2

Query Match 6.2%; Score 36.6; DB 1; Length 913;
Best Local Similarity 58.9%; Pred. No. 0.21; Indels 0;
Matches 63; Conservative 0; Mismatches 44; Gaps 0;
QY 198 TCATACCAATAAGGACTCTGGCGTGTGCATCCCCCTGCAACCCCTGTTACGGCC 257
Db 320 TCCACCCAGTTCCTCCTCCCTGCAACTCCACCCCTGCTTCCTCTGCAACTCC 379
QY 258 AGCCAAGGCCAGCGCGCTCGCCGGGGGATCGCTAGGCC 304
Db 380 ACCTCCAGCTCTCCACCTCTGCCACTCTCCACCAAGCTCTCCAC 426

RESULT 9
US-08-787-335-2
; Sequence 3, Application US/08298829
; Patent No. 5620882

GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,829
; FILING DATE: 19-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/885,970
; FILING DATE: 18-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
; US-08-298-829-3

Query Match 6.2%; Score 36.6; DB 1; Length 913;
Best Local Similarity 58.9%; Pred. No. 0.21; Indels 0;
Matches 63; Conservative 0; Mismatches 44; Gaps 0;
QY 198 TCATACCAATAAGGACTCTGGCGTGTGCATCCCCCTGCAACCCCTGTTACGGCC 257
Db 320 TCCACCCAGTTCCTCCTCCCTGCAACTCCACCCCTGCTTCCTCTGCAACTCC 379
QY 258 AGCCAAGGCCAGCGCGCTCGCCGGGGGATCGCTAGGCC 304
Db 380 ACCTCCAGCTCTCCACCTCTGCCACTCTCCACCAAGCTCTCCAC 426

RESULT 10
US-08-787-335-2
; Sequence 2, Application US/08787335
; Patent No. 5981834

STATE: Wisconsin
 COUNTRY: USA
 ZIP: 53701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Microsoft Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/298, 687A
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/617, 239
 FILING DATE: 21-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/253, 243
 FILING DATE: 04-OCT-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27,396
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 283-2478
 TELEFAX: (608) 251-5139
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1985 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 LENGTH: 1985 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 LENGTH: 1985 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Gossypium barbadense
 STRAIN: Sea Island
 IMMEDIATE SOURCE:
 LIBRARY: EMBL-SI
 CLONE: S1H6
 US-08-298-687A-25

Query Match 6.2%; Score 36.6; DB 1; Length 1985;
 Best Local Similarity 58.9%; Pred. No. 0.27;
 Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 TCAATACCAATAATCGACTTCTGCGCTGTCACTCCCTGGATACCCCTGTAGGCCTAGGCC 257
 DB 569 TCCACCCAGTTCTCTCTCTGCAACTCCACCACTCCTCTCTGCAACTCC 628
 QY 258 AGCCAAAGGCAGCCGGCTCTGGCGGGATGGCTGAGCGC 304
 DB 629 ACCTCCAGCTTCACCTCCACCTCCACGTTCTCCAC 675

RESULT 13
 US-08-298-829-25

Sequence 25, Application US/08298829
 Patent No. 5620882

GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON

NUMBER OF SEQUENCES: 33

TITLE OF INVENTION: PLANTS FOR ALTERED FIBER

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nicholas J. Seay, Quarles & Brady
 STREET: P.O. Box 2113, First Wisconsin Plaza
 CITY: Madison
 STATE: Wisconsin

COUNTRY: USA

ZIP: 53701

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Microsoft Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/298, 829
 FILING DATE: 19-OCT-1994
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/885, 970
 FILING DATE: 18-MAY-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/617, 239
 FILING DATE: 21-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/253, 243
 FILING DATE: 04-OCT-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27,386
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 283-2478
 TELEFAX: (608) 251-139
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1985 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: Linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Gossypium barbadense
 STRAIN: Sea Island
 IMMEDIATE SOURCE:
 LIBRARY: EMBL-SI
 CLONE: S1H6
 US-08-298-829-25

Query Match 6.2%; Score 36.6; DB 1; Length 1985;
 Best Local Similarity 58.9%; Pred. No. 0.27;
 Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 TCATACCATACGACTTCTGCGCTGTCACTCCCTGGATACCCCTGTAGGCC 257
 DB 569 TCCACCCAGTTCTCTCTGCAACTCCACCACTCCTCTGCAACTCC 628
 QY 258 AGCCAAAGGCAGCCGGCTCTGGCGGGATGGCTGAGCGC 304
 DB 629 ACCTCCAGCTTCACCTCCACCTCCACGTTCTCCAC 675

RESULT 14
 US-08-298-114-1

Sequence 1, Application US/09130114
 Patent No. 5976307

GENERAL INFORMATION:

APPLICANT: Horlick, Robert A.
 APPLICANT: Damai, Bassam B.
 APPLICANT: Robbins, Alan K.
 TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes From Multiple Transfected Episomes
 FILE REFERENCE: 086/1D90JUS1
 CURRENT APPLICATION NUMBER: US/09/130,114
 CURRENT FILING DATE: 1998-08-06
 NUMBER OF SEQ ID NOS: 36
 SOFTWARE: FastSEQ for Windows Version 3.0
 SEQ ID NO 1
 LENGTH: 5452
 TYPE: DNA
 ORGANISM: VENNA
 US-09-130-114-1

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 19:13:05 ; Search time 1279.51 Seconds
(without alignments)
1752.815 Million cell updates/sec

Title: US-08-988-242-1_COPY_1232_1825
Sequence: 594 1 CAGGTACAGCTAACGGCTT.....AAACGAGCTCGTGGCGATA 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*

1: em_est1:
2: em_est2:
3: em_est3:
4: em_est4:
5: em_est5:
6: em_est6:
7: em_est7:
8: em_est8:
9: em_est9:
10: em_est10:
11: em_est11:
12: em_est12:
13: em_est13:
14: em_est14:
15: em_est15:
16: em_est16:
17: em_est17:
18: em_est18:
19: em_est19:
20: gb_est1:
21: gb_est2:
22: gb_est3:
23: gb_est4:
24: gb_est5:
25: gb_est6:
26: gb_est7:
27: gb_est8:
28: gb_est9:
29: gb_est10:
30: gb_est11:
31: gb_est12:
32: gb_est13:
33: gb_est14:
34: gb_est15:
35: gb_est16:
36: gb_est17:
37: gb_est18:
38: gb_est19:
39: gb_est20:
40: gb_est21:
41: gb_est22:
42: gb_est23:
43: gb_est24:
44: gb_est25:

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match Length	DB ID	Description
c 1	47.8	8 0	925	79	CNS0091P
c 2	43.8	7 4	1036	79	CNS010BS

45: gb_est26:
46: gb_est27:
47: gb_est28:
48: gb_est29:
49: gb_est30:
50: gb_est31:
51: gb_est32:
52: em_est20:
53: em_est21:
54: em_est22:
55: em_est23:
56: em_est24:
57: em_est25:
58: em_est26:
59: em_est27:
60: em_est28:
61: em_est29:
62: em_est30:
63: em_est31:
64: em_est32:
65: em_est33:
66: em_est34:
67: em_est35:
68: em_est36:
69: em_est37:
70: em_est38:
71: em_est39:
72: em_est40:
73: em_est41:
74: em_est42:
75: em_est43:
76: em.est32:
77: em.est33:
78: em.est34:
79: gb.gss1:
80: gb.gss2:
81: gb.gss3:
82: gb.gss4:
83: em.gss1:
84: em.gss2:
85: em.gss3:
86: em.gss4:
87: gb.gss5:
88: gb.gss6:
89: gb.gss7:
90: gb.gss8:
91: gb.gss9:
92: em.gss4:
93: em.gss5:
94: em.gss6:
95: em.gss7:
96: em.gss9:
97: em.gss10:
98: em.gss11:
99: gb.gss10:
100: gb.gss11:
101: em.gss12:
102: gb.gss12:
103: gb.gss13:
104: gb.gss14:
105: gb.gss15:

SUMMARIES

Result No.	Query No.	Score	Match Length	DB ID	Description
c 1	47.8	8 0	925	79	CNS0091P
c 2	43.8	7 4	1036	79	CNS010BS

AL053013 Drosophil
AL098770 Drosophil

library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Geneviève Payan. It has been constructed in the vector pBeloBAC11.

FEATURES
source
1. .1036
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03J011"
/note="end : SP6"
240 a 127 c 155 g 234 t 280 others
BASE COUNT
ORIGIN

Query Match 7.4%; Score 43.8; DB 79; Length 1036;
Best Local Similarity 23.6%; Pred. No. 2.5; Mismatches 113; Indels 0; Gaps 0;
Matches 92; Conservative 113; Mismatches 185; Indels 0; Gaps 0;

QY 94 CGGTCTCCGAGGGAAAGAACAGGCCAAACATGGTAGTGGGACGGC 153
Db 637 CKACCSSSAVMSASARAAAAYAASAAAVAGMARAAGASCGRSSSCRG 696

QY 154 AACCGGGTTGTTGTCCTGGGACTGACGGGAGTAGCAGTCATAACATTAGCT 213
Db 697 CSVCYCKCGCCCYGAANGAACGAGSAAATGCGSAGCAGGCAAGMMGASCSAASA 756

QY 214 TCTGCCTGCTGTCATCCCCCTGCATCACCCCCCTGCTGGCCAGGCCAGGG 273
Db 757 RAVCGCGCSGSCGVSCASCSGCGNSCGAMASSSSCASWAGASASGSGCC 816

QY 274 CTCCTTGCCGGCGCATGGCTGAGCCAGCTGGGAGCAGATCATTTGCTTAATCTA 333
Db 817 SCSCGASSCVSGSCGVSCGCGMCVSCGCGNSCGCASSGGSASASS 876

QY 334 GTGAATCAGCTGGATTAAATGTCACCAAAAGGACGGCTGCAAGCTGGCCGGCC 393
Db 877 SGSTCGSSSSGGGASMGSSSGGSSVARGCSARAGCRRASSGGCSCACNSC 936

QY 394 ACGAGCAGGTCTAGCGGGTAGCTGGCAGCTACGCCGCCGACAGTCATAC 453
Db 937 ASMGCKCGMAAASVSGMCAAAAGGACGSSSSCGSGSVSBSTCGSARVAR 996

QY 454 GGGCACAAATGGCGACCTGAGCGCTGGA 483
Db 997 SGGGCAGGSSGSSGSSVSSVSGSSGGGA 1026

RESULT 3
CNS0108S CNS0108S 1065 bp DNA GSS 26-JUL-1999
LOCUS DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN03J03 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098662.1 GI:5610273
VERSION GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Pterygota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Muscomorpha; Ephemeroidea; Diptera; Brachycera;
Muscidae; Drosophilidae; Drosophila.

REFERENCE 3
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Séquençage : BP 191 91006 EVRY Cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determinatoin of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Geneviève Payan. It has been constructed in the vector pBeloBAC11.

FEATURES
source
1. .1065
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03J03"
/note="end : T7"
252 g 313 t 154 others
BASE COUNT
ORIGIN

Query Match 7.3%; Score 43.4; DB 79; Length 1065;
Best Local Similarity 27.4%; Pred. No. 3.1; Mismatches 62; Indels 73; Gaps 0;
Matches 51; Conservative 51; Mismatches 73; Indels 0; Gaps 0;

QY 141 AGTGGGAGGGAAACCGGGTGTGTCCTGGGACTGACGGGGAGTAGCAGTCA 200
Db 755 RVGCGSSSSGGSSGSSGSSCTUSTKGKGGSSSSKGSSSSKBSSBKSSEGGKT 814

QY 201 TACCAATAAGCAGCTGCGCTGCTGATCCCTGATCACCCCTGTGTCAGGCCAGC 260
Db 815 CACCTTBBGTTKGTKTNGKTRTCSCCGGYYTCCCCTCCCGSCGNCMCCCCSC 874

QY 261 CAAGGAGCCGCGCTCTCGCGCCGATCGCTGAGCCGACGTGGGAGCAAGAT 320
Db 875 CMTGCCTTSSSSSSSSSSSSCTKCCSKTGSSSGGSSCTRKGVK 934

RESULT 4
AA858891 AA858891/c mRNA EST 03-JUL-1999
LOCUS UI-R-A0 bd-d-01-0-UI-s1 UI-R-A0 Rattus norvegicus CDNA clone
DEFINITION UI-R-A0 bd-d-01-0-UI 3, similar to gb|AA422729|AA422729|vd29p09.s1
knowles Solter mouse 2 cell Mus musculus CDNA clone 793913 5 , mRNAs
sequence.
ACCESSION AA858891
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Butherus; Rodentia; Sciurognath; Muridae; Murinae; Rattus.
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948242.
REFERENCE 1 (bases 1 to 447)
AUTHORS Bonaldo, M.F.; Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery.
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948242.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dt track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cdna library preparation: M.
Patima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics. This clone is also available through the
I.M.A.G.E. Consortium at LNL (infoimage.lnl.gov). IMAGE
ID=176747 The following repetitive elements were found in this
CDNA sequence: 15-121, >(CA)n#Simple_repeat 61-179,

Scoring table:	IDENTITY_NUC				
Searched:	Gapext 1.0	Gapext 10.0 , Gapext 1.0			
Total number of hits satisfying chosen parameters:	31185 seqs, 125096042 residues	623170			
Minimum DB seq length:	0				
Maximum DB seq length:	10000000				
Post-processing:	Minimum Match 0%	Listing first 45 summaries			
Database :	N_Geneseq_36-*				
Pred.	No.	is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
		SUMMARIES			
Result No.	Score	Query Match	Length	DB ID	Description
1	590.8	99.5	3402	T27310	Trypanosoma cruzi
2	39.2	6.6	2790	V18480	BOP1 cDNA. New iso
C	38.2	6.4	2943	T16480	SP8 virus 9B glyco
C	36.8	6.2	3776	T42903	TRP-1 protein codi
C	36.6	6.2	203	V17226	SCA2 gene CAG repe
C	36.6	6.2	203	V30271	Glutamine rich reg
C	36.6	6.2	913	T07199	Cotton fiber-speci
C	36.6	6.2	913	T1034	Cotton fibre-speci
C	36.6	6.2	913	T30253	Cotton fibre speci
C	36.6	6.2	913	T65610	Cotton fibre speci
C	36.6	6.2	913	T70041	Cotton fibre speci
C	36.6	6.2	1984	T13030	Cotton fibre clone
C	36.6	6.2	1985	T30050	Cotton H6 gene and
C	36.4	6.1	1985	T70036	cDNA of the M3/6 g
C	36.4	6.1	2415	T86757	cDNA of the M3/6 g
C	16	6.1	2453	T86758	Genomic Eimeria te
C	17	6.1	633	Q03122	F1GA insert stabil
C	18	6.1	795	V55830	Nucleotide sequenc
C	19	6.1	799	V55831	Vector plasmid pCM
C	20	6.1	9600	V21683	Plasmid pCISBON f
C	21	36	6.1	10596	Plasmid pCISBON f
C	22	36	6.1	10596	Nucleotide sequenc
C	23	36	6.1	10596	Receptor type tyro
C	24	35.8	6.0	3076	Sequence of the I.
C	25	35.6	6.0	V43674	Eimeria tenella sp
C	26	35.6	6.0	N92576	Human adenosine AI
C	27	35.6	6.0	1089	Human secreted pro
C	28	35.4	6.0	T93553	Antigen tc-11 gen
C	29	35.2	5.9	X53955	AF-9 cDNA. New acu
C	30	35.2	5.9	V61463	Spinocerebellar at
C	31	35	5.9	V61463	Spinocerebellar at
C	32	35	5.9	Q23081	Spinocerebellar at
C	33	35	5.9	Q23082	Spinocerebellar at
C	34	35	5.9	Q84832	Spinocerebellar at

Query	Match	6 2%	Score	36.8	DB	1;	Length	3776;
Best Local Similarity	69.4%	Pred.	No. 1.4;					
Matches	50;	Conservative	0;	Mismatches	22;	Indels	0;	Gs
Qy	231	CCCTGATACCCCTGTTCAGGCCAAGCAGCCCTCTGGCCGGCG						
Db	1355	CCCCGCACGCCAGCCCCAGGCCAGCCACCGCAGCCAGCTGAGTCAGGCC						
Query	291	ATGGCTGAGCC	302					
Db	1415	AGCCCCAGAGCC	1426					

RESULT	5
V17226/c	
ID	V17226 standard; DNA; 203 BP.
AC	V17226;
DT	29-JUN-1998 (first entry)
DE	SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
KW	SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
Synthetic.	
OS	
WO	WO9803679-A1.
PN	
PD	29-JAN-1991.
PF	18-JUL-1996; JO1999.
PR	18-JUL-1996; WO-JO1999.
PA	(SRLS) SRL INC.
PI	Sanpei K., Tsuji S.
DR	WPI: 98-120706/11.
PT	Diagnosing spinocerebellar ataxis type II - by PCR and determining

This sequence represents a fragment of the SCA2 gene. It can be used in the method of the invention for diagnosing spinocerebellar ataxia type II, by performing PCR on the test DNA using two primers hybridizing to parts of the SCA2 gene sequence, and determining the number of CAG repeats in the amplified products. The method provides an easy means for diagnosis of spinocerebellar ataxia type II.

Example 1; Page 22; 38pp; Japanese.
 This represents a sequence encoding glutamine rich repeat region which can be found in spinocerebellar atrophy type 2 (SCA2) patients. The specification provides a gene sequence causative of the neurodegenerative disease SCA2, having a tri-nucleotide (CAG) repeat region which in the expression product produces a polyglutamine sequence from Gin-16 to Gin-188. In the normal gene there are 15-25 CAG repeats but in SCA2 patients this number is increased to 35-100. Peptides encoded by nucleic acid fragments (DNA or RNA) containing sequences from the SCA2 associated gene, antibodies recognising the peptides and antisense nucleic acids hybridising with the nucleic acid fragments can be used for the investigation and diagnosis of SCA2. They can also be used for the treatment of SCA2 by antisense therapy or gene therapy.

Sequence	203 BP;	68 A;	70 C;	62 G;	3 T;
----------	---------	-------	-------	-------	------

Query Match	Best Local Similarity	Score	DB 1;	Length	203;
	Matches	66;	Conservative	0;	Gaps
			Mismatches	49;	Indels
N	184	GCGCGAGTAGCAGTCATACCAATACTACGACTTCCGCGTGTGCATCCCTGCATCACCC	243		0
O	202	GCCAGAATTTCCGTGATGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	143		0
Y	244	CCTGTTTCAGGCCAGGCCAAGGCCAGCCGCCCTCTGGCGGGCATGGCT	298		0
D	142	GGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	88		0

RESULT 7

PT	New isolated fibre-specific promoters - used for introducing altered fibre-specific characteristics into plants, partic. cotton.
ID	T07199 standard; DNA; 913 BP.
AC	T07199; 10-MAY-1996 (first entry)
DT	Cotton fiber-specific H6 protein gene. Cotton; transgenic plant; fiber; immobilisation; enzyme; pesticide degradation; parathion hydrolase; extensin; ds.
DE	H6 gene; promoter; cotton; transgenic plant; fiber; immobilisation; enzyme; pesticide degradation; parathion hydrolase; extensin; ds.
KW	Gossypium hirsutum strain Coker 312.
OS	Key
OS	Location/Qualifiers
FT	72. 716
FT	signal_peptide / [*] tag= a
FT	mat_peptide / [*] tag= b
FT	148. 713
FT	/ [*] tag= c
PN	US5474925-A.
PD	12-DEC-1995.
PF	19-DIC-1991; 812233.
PR	19-DIC-1991; US-812233.
PR	23-MAR-1994; US-217327.
PR	(CETU) AGRACTETUS INC.
PI	Barton KA, Maliyakal J;
WPI	96-019513/04.
P-PSDB	R86913.
PT	Prep. of immobilised enzymes in cotton fibre - esp. pesticide degrading enzymes e.g. parathion hydrolase.
PT	Disclosure: Column 31-34; 37 pp; English.
CC	The cotton H6 gene (T07198) encodes an extensin-like protein (R86913) that is an integral part of the secondary cell wall of cotton. The portions of the gene encoding the N-terminal or transmembrane regions of the protein may be utilised in the construction of gene fusions used to anchor (i.e. immobilise) foreign proteins in the fibers of transgenic cotton plants. Fusions of H6 and a foreign protein should be localised within the secondary walls of the fiber. The immobilised proteins, e.g. enzymes or antibodies, have industrial, medical and environmental applnns. Sequence 913 BP.; 196 A; 313 C; 155 G; 249 T;
CC	Sequence 913 BP.; 196 A; 313 C; 155 G; 249 T;
CC	REUSLT 8
CC	T13034 standard; cdNA; 913 BP.
CC	ID T13034; 27-MAY-1995 (first entry)
CC	Cotton fibre-specific cdNA clone H6. Cotton; fibre; promoter; transgenic plant; crop improvement; ds.
CC	KW Location/Qualifiers
CC	Key
CC	72. 716
CC	/ [*] tag= a
CC	US5495070-A.
CC	PD 27-FEB-1996.
CC	PF 04-OCT-1988; 253243.
CC	PR 04-OCT-1988; US-553243.
CC	PR 21-NOV-1990; US-617239.
CC	PR 18-MAY-1992; US-655970.
CC	PR 19-OCT-1994; US-298687.
CC	PA (CETU) AGRACETUS INC.
CC	John M; WPI: 96-267704/27.
CC	PT Isolation of fibre-specific cotton promoter sequences - using selected DNA probes to screen genomic DNA fragments, for production of cotton fibres with improved characteristics.
CC	AC Cotton fibre cell-specific promoter sequences were isolated by differential screening of a cotton plant cdNA library. Of 4788 clones from 10 day cell library screened with leaf cdnas, 800 clones not present in the leaf were isolated. These were screened with cdnas from ovule, root and flower mRNAs and resulted in 79 clones isolated. PCR analysis was then used to remove cross hybridising clones. This resulted in the isolation of 18 cdna clones specifically expressed in cotton fibres (T30242-4 and T30253-57). These cdnas were then used to screen for homologous genomic sequences (T30245-53 and T30268) in order to obtain the corresp. promoter sequences.
CC	This cdna clone contains an insert of 500 bp which hybridised with a 950 bp RNA. It is expressed in fibres but not in root, flower, leaf or ovules. The sequence suggests a single open reading frame extending from bases 71-710. This sequence has been deposited with ATCC at accession no. 67810.
CC	The promoters isolated from the fibre cell-specific clones can be used to generate transgenic cotton plants and lines producing fibres having altered quantity and quality.
CC	Sequence 913 BP.; 196 A; 313 C; 155 G; 249 T;

RESULT		4	
Qy	389	CGGCCACGAGGAGGTCTAAGGGTGTACGGGGTCAAGACTACGCCGCAGCGAC 4.43	T4 2903
Db	75	CGGCCGCTAGGGCAGACTACGGGCCGTGACTCCGGCCGTCAGCTCGGCCGCGCGGC 21	ID AC DT DE KW KW KW KW KW KW KW Homo OS FH FT FT FT PN PD PF PR PR PA PA DR DR PT PT PT PT PS CC CC CC CC CC CC CC CC CC SO
			T4 2903 standard; DNA: 3776 BP. T4 2903; 16-JUN-1997 (first entry) TRP-1 protein coding sequence. Gene name; regulation; plasmid; viral infection; human T-cell leukaemia; HIV; antiviral agent; detection; cancer; gene therapy; TSP; ds. Homo sapiens. Location/Qualifiers FT /*tag= a /product= TRP-1 WO9630522-A1. 03-OCT-1996. 19-MAR-1996; J00719. 24-MAR-1995; JP-065559. 27-APR-1995; JP-104299. (SHIO) SHIRONOGI & CO LTD. Igarashi H, Okumura K, Orita S, Saiga A, Sakaguchi G; WPI: 96-5536745. P-PDB; W06136. DNA molecule with gene expression regulation activity - for use in treatment of human T-cell leukaemia and HIV, as antivirus agent and for detecting cancer Claim 17: Page 48-54; 77pp; Japanese. This sequence encodes the protein TRP-1. This sequence was used in conjunction with a DNA molecule with gene expression regulation activity for regulation of gene expression, and treatment of infection pref. human T-cell leukaemia and HIV. The plasmid also a protein which is used as an antivirus agent, and also in a method detecting cancer. The DNA molecule and protein have potential use in gene therapy, and the plasmid may also have potential use in the treatment of TSP. Sequence 3776 BP; 764 A; 1177 C; 1183 G; 652 T;
Query Match	6.2%	Score 36.8; DB 1; Length 3776;	
Best Local Similarity	69.4%	Pred. No. 1.4;	
Matches	50;	Mismatches 0; Indels 0;	G
Qy	231	CCCTGGATCCCCCTGTTTACGGCCAGGCCAAGGCCAGCGCCCTCCCTGCCGGCGCG	
Db	1355	CCCCGCACGCCAGCCCACGCCAGCGCAGCTGAGTCAGCTGCAGGCC	
Qy	291	ATCGGCTGACCC 302	
Db	1415	AGCCCCAGACCC 1426	

RESULT	5
V17226/C	
ID	v17226 standard; DNA; 203 BP.
AC	V17226
DT	29-JUN-1998 (first entry)
DE	SCA2 gene: spinocerebellar ataxis type II; CAG repeat; PCR primer
KW	
OS	Synthetic.
PN	W09803679-A1.
PD	29-JAN-1998.
PF	18-JUL-1996; J01999.
PR	18-JUL-1996; WO J01999.
PA	(SRLS) SRL INC.
PI	Sanpei K, Tsuji S;
WPI	98-120796/11.
DR	Diagnosing spinocerebellar ataxis type II - by PCR and determinin
PT	

Query Match	Match	Score	DB	Length	0;
Best Local Similarity	6.2%	36.6	1;	203;	
Best Local Conservative Matches	57.4%	Prod. No.	0.79;		
Best Local Mismatches	66	0;	Mismatches	49;	Indels 0; Gaps 0;
Y	184	GGGGCGAGTACGAGTCATACCAATACTGGACTTTGGCCGCTGTGCATCCCCCTGCATCACCC	243		
O	202	GCCCAGACTTCGGTGTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT	143		
Y	244	CCGGTTTACGGCCAGCAAGGAGCGCCCTCCTGGCGGCGATCGCGCTG	298		

RESULT 7

T07199	standard; DNA; 913 BP.	PT New isolated fibre-specific promoters - used for introducing altered fibre-specific characteristics into plants, partic. cotton.
AC T07199		PT Example 3; Column 25-28; 46pp; English.
DT 10-MAY-1996 (first entry)		PS Cotton fibre-specific cDNA clone H6 (T13034) (ATCC 67810) was isolated from a cDNA library of cotton var. Coker 312 15-day-old boll samples by subtractive hybridization procedure. It hybridizes to a developmentally regulated RNA of 950 bases that is not detected in leaf, flower, ovule or root. The putative protein product is proline rich and clearly distinct from previously known plant cell wall proteins. H6 and other fibre-specific cDNA clones (see T13033-T13050) were used to screen cotton genomic libraries, leading to the isolation of genomic clones (see T13025-32 and T13052-53) contg. sequences capable of promoting gene expression in fibre cells.
Key		CC
Location/Qualifiers		CC
72..716		CC
signal_peptide /*tag= a		CC
72..147		CC
/*tag= b		CC
mat_peptide /*tag= b		CC
148..713		CC
/*tag= C		CC
PT US5474925 A.		CC
PD 12-DEC-1995.		CC
PF 19-DEC-1991; 812233.		CC
PR 19-DEC-1991; US-811233.		CC
PR 23-MAR-1994; US-211327.		CC
PA (CETU) AGRACETUS INC.		CC
PI Barton KA, Malaiyakal J;		CC
WPI; 96-039513/04.		CC
DR P-PSDB; R86913.		CC
PT Prepn. of immobilised enzymes in cotton fibre - esp. pesticid-degrading enzymes e.g. Enzymatic hydrolyase		CC
PS Disclosure: Column 31-34; 37pp; English.		CC
CC The cotton H6 gene (T07199) encodes an extensin-like protein (R8613) that is an integral part of the secondary cell wall of cotton. The portions of the gene encoding the N-terminal or transmembrane regions of the protein may be utilised in the construction of gene fusions used to anchor (i.e. immobilise) foreign proteins in the fibers of transgenic cotton plants.		CC
CC Fusions of H6 and a foreign protein should be localised within the secondary walls of the fiber. The immobilised proteins, e.g. enzymes or antibodies, have industrial, medical and environmental applns.		CC
Sequence 913 BP.; 196 A; 313 C; 155 G; 249 T;		CC
PT US5474925 A.		CC
PD 12-DEC-1995.		CC
PF 19-DEC-1991; 812233.		CC
PR 19-DEC-1991; US-811233.		CC
PR 23-MAR-1994; US-211327.		CC
PA (CETU) AGRACETUS INC.		CC
WPI; 96-039513/04.		CC
DR P-PSDB; R86913.		CC
PT Prepn. of immobilised enzymes in cotton fibre - esp. pesticid-degrading enzymes e.g. Enzymatic hydrolyase		CC
PS Disclosure: Column 31-34; 37pp; English.		CC
CC The cotton H6 gene (T07199) encodes an extensin-like protein (R8613) that is an integral part of the secondary cell wall of cotton. The portions of the gene encoding the N-terminal or transmembrane regions of the protein may be utilised in the construction of gene fusions used to anchor (i.e. immobilise) foreign proteins in the fibers of transgenic cotton plants.		CC
CC Fusions of H6 and a foreign protein should be localised within the secondary walls of the fiber. The immobilised proteins, e.g. enzymes or antibodies, have industrial, medical and environmental applns.		CC
Sequence 913 BP.; 196 A; 313 C; 155 G; 249 T;		CC
Query Match 6.2%; Score 36.6; DB 1; Length 913;		CC
Best Local Similarity 58.9%; Pred. No. 1.1;		CC
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;		CC
QY 198 TCATACCAATAGCATCTTCGGCTGCTCATCCCCGTGATCACCCCTGTTTCAGGCC 257		CC
DB 320 TCCACCCAGTTCTCCCTCTGAACTCACCACTCCTCTCCACTCC 379		CC
QY 258 AGCCAAAGGCAGCGCGCCMCTGGCGGGATGGCTGAGCCGC 304		CC
DB 380 ACCCTCCAGTTCTACCTTCGGCTGACCTCCCTCCAGTTCTCCAC 426		CC
RESULT 8		CC
AC T13034 standard; cDNA; 913 BP.		CC
DT 27-MAY-1996 (first entry)		CC
DE Cotton fibre-specific cDNA clone H6.		CC
KW fibre promoter; transgenic plant; crop improvement; ds.		CC
OS Gossypium hirsutum strain Coker 312.		CC
FH Location/Qualifiers		CC
72..716		CC
/*tag= a		CC
PT US5495070 A.		CC
PD 27-FEB-1996.		CC
PF 04-OCT-1988; 253243.		CC
PR 04-OCT-1988; US-253243.		CC
PR 21-Nov-1990; US-617239.		CC
PR 18-MAY-1992; US-885970.		CC
PA (CETU) AGRACETUS INC.		CC
WPI; 96-2677794/27.		CC
PT Isolation of fibre-specific cotton promoter sequences - using selected DNA probes to screen genomic DNA fragments, for production of cotton fibres with improved characteristics.		CC
PT Examples; Column 25-28; 46pp; English.		CC
PS Cotton fibre cell-specific promoter sequences were isolated by differential screening of a cotton plant cDNA library. Of 4788 clones from a 10 day cell library screened with leaf cDNAs, 800 clones not present in the leaf were isolated. These were screened with cotton, root and flower mRNAs and resulted in 79 clones isolated. PCR analysis was then used to remove cross-hybridising clones. This resulted in the isolation of 18 cDNA clones specifically expressed in cotton fibre cells (T30242-4 and T30253-67). These cDNAs were then used for homologous genomic sequences (T30245-53 and T30268) in order to obtain the corresp. promoter sequences.		CC
CC This cDNA clone contains an insert of 500 bp which hybridised with a 950 bp RNA. It is expressed in fibres but not in root, flower, leaf or ovules. The sequence suggests a single open reading frame extending from bases 71-710. This sequence has been deposited with ATCC at accession no. 67810.		CC
CC The promoters isolated from the fibre cell-specific clones can be used to generate transgenic cotton plants and lines producing fibres having altered quantity and quality.		CC
CC sequence 913 BP.; 196 A; 313 C; 155 G; 249 T;		CC
DR WPI; 96-139095/14.		CC

Query Match 6.2%; Score 36.6; DB 1; Length 913;
 Best Local Similarity 58.9%; Pred. No. 1.1; Indels 0; Gaps 0;
 Matches 63; Conservative 0; Mismatches 44; Delins 0; Gaps 0;

Oy 198 TCATACCAATATGAACTTCTCGTGTGCATCCCCCTGATCACCCCTGTTCAGGCC 257
 Db 320 TCCACCCAGTCTCTCCCTCCTGAACTCCACCTCCACTCTCCCTCCACTCCAACTCC 379

Oy 258 AGCCAAAGGCGAGGGCGGCGCTCCGGGGGGATGGCTGAGGCC 304
 Db 380 ACCTCCAGGTCTCCACCTCTGCCACTCTCCACCCAGCTTCTCAC 426

RESULT 10

T26210 standard; cDNA to mRNA; 913 BP.
 AC T62611;
 DT 14-MAY-1997 (first entry)
 DE Cotton fibre specific cDNA clone CKFB15A1-H6.
 KW cotton; fibre-specific; strength; transgenic plant; anthesis;
 RW Gossypium hirsutum strain Coker 312; H6; antisense; sense; ss.
 OS US5597718-A.
 PN PR 04-OCT-1988; 253243.
 PD 28-NOV-1997.
 PR 04-OCT-1988; US-617239.
 PR 18-OCT-1993; US-138814.
 PR 20-SEP-1995; US-530797.
 PA (CETU) AGRACETUS.
 PI Brill WJ, John ME, Umbeck PF;
 DR WPI; 97-108326/10.
 PR Prodn. of transgenic cotton plants - by transformation with the H6 coding sequence or E6 anti-sense sequence, produces fibre of altered strength.
 PR Claim 1; Column 29-30; 33bp; English.
 PS T62609-24 are cotton fibre-specific cDNA clones which can be used to identify genomic clones. This clone, CKFB15A1-H6, is developmentally regulated. The H6 RNA was found to be developmentally regulated. (CK = Coker; FB = Fibre; 10, 15 or 23 = age in days of fibre cells; A1 and the last character and number stand for clone identity). The fibre-specific genes were identified by differential cDNA library screenings. Coding sequences from these isolated genes are used in sense or antisense orientation to alter the fibre characteristics, e.g. strength, of transgenic fibre producing plants.
 SQ Sequence 913 BP; 196 A; 313 C; 155 G; 249 T; 249 T;

Query Match 6.2%; Score 36.6; DB 1; Length 913;
 Best Local Similarity 58.9%; Pred. No. 1.1; Indels 0; Gaps 0;
 Matches 63; Conservative 0; Mismatches 44; Delins 0; Gaps 0;

Oy 198 TCATACCAATATGAACTTCTCGTGTGCATCCCCCTGATCACCCCTGTTCAGGCC 257
 Db 320 TCCACCCAGTCTCTCCCTCCTGAACTCCACCTCCACTCTCCCTCCACTCCAACTCC 379

Oy 258 AGCCAAAGGCGAGGGCGGCGCTCCGGGGGGATGGCTGAGGCC 304
 Db 380 ACCTCCAGGTCTCCACCTCTGCCACTCTCCACAGCTTCTCAC 426

RESULT 11

T70041 standard; cDNA; 913 BP.
 AC T70041;
 DT 20-AUG-1997 (first entry)
 DE Cotton fibre specific cDNA clone H6.
 KW heterologous gene expression; ds.
 OS Gossypium hirsutum strain Coker 312.
 FH Key
 FT 71 .710 /*tag= a

Query Match 6.2%; Score 36.6; DB 1; Length 913;
 Best Local Similarity 58.9%; Pred. No. 1.1; Indels 0; Gaps 0;
 Matches 63; Conservative 0; Mismatches 44; Delins 0; Gaps 0;

Oy 198 TCATACCAATATGAACTTCTCGTGTGCATCCCCCTGATCACCCCTGTTCAGGCC 257
 Db 320 TCCACCCAGTCTCTCCCTCCTGAACTCCACCTCCACTCTCCCTCCACTCCAACTCC 379

Oy 258 AGCCAAAGGCGAGGGCGGCGCTCCGGGGGGATGGCTGAGGCC 304
 Db 380 ACCTCCAGGTCTCCACCTCTGCCACTCTCCACAGCTTCTCAC 426

RESULT 12

T13030 standard; DNA; 1984 BP.
 ID T13030;
 AC T13030;
 DT 27-MAY-1996 (first entry)
 DE Cotton fibre-specific H6 gene and promoter; ds.
 KW Cotton; fibre; promoter; transgenic plant; crop improvement; ds.
 OS Gossypium barbadense strain Sea Island.
 FH Key
 FT Promoter
 FT /*tag= a
 FT 3'utr
 FT /*tag= b
 FT cds
 FT /*tag= c
 FT /*tag= "the CDS includes a 583 bp intron
 FT (location unspecified)"
 FT /*tag= d
 PR 1548 .1984
 PR 253243
 PR 04-OCT-1988; US-253243.
 PR 21-NOV-1990; US-617239.
 PR 18-MAY-1992; US-885970.
 PA (CETU) AGRACETUS INC.

PI John M;
 DR WPI; 96-139095/14.
 PT New isolated fibre-specific promoters - used for introducing
 altered fibre-specific characteristics into plants, partic. cotton.
 PS Claim 1; Column 59-62; 48pp; English.
 CC The cotton var. Sea Island H6 gene and promoter (T13030) were
 identified in genomic clone EMBL-SI-H6-4 isolated by screening a
 genomic library prepd. in lambda EMBL with cotton fibre-specific
 DNA clone H6 (see T13034). H6 and other (see T13025-29, T1301-32
 and T13052-53) fibre-specific promoters can be utilised in the
 construction of vectors used for the introduction of altered fibre-
 specific characteristics into plants, partic. cotton. They can be
 used to modulate the synthesis of fibre proteins or to introduce
 non-fibre proteins into fibre in a tissue-specific manner.
 SO Sequence 1984 BP; 565 A; 470 C; 281 G; 668 T;

Query Match 6.2%; Score 36.6; DB 1; Length 1984;
 Best Local Similarity 58.9%; Pred. No. 1..3;
 Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 TCATACCAATTAGACTCTCGCGTCGATCCCCCTGATACCGCTGTTCAGGCC 257
 DB 569 TCCACCCAGTTCTCCTCTCTGAACTCCACCTGCTTCCTCCGAACTCC 628

QY 258 AGCCAAAGCAGCGCGCCGGCTGGCGCATGGCTGAGCGCG 304
 DB 629 ACCTCCAGCTCTCCACCTGCTCCACCTCCACCGCTTCTCCAC 675

RESULT 14
 T70036 standard; DNA; 1985 BP.
 ID T70036; standard; DNA; 1985 BP.
 AC T70036;
 DT 20-AUG-1997 (first entry)
 DE Cotton H6 gene and fibre-specific promoter from clone SIH6.
 KW cotton; E6; fibre; promoter; transgenic plant; truncated;
 OS Gossypium barbadense strain Sea Island.
 PN US5620882-A.
 PR 04-OCT-1988; 253243.
 PD 15-APR-1997.
 PR 04-OCT-1988; US-253243.
 PR 21-NOV-1990; US-617239.
 PR 18-MAY-1992; US-885970.
 PR 19-OCT-1994; US-298829.
 PA (CETU) AGRACETUS INC.
 PI John M;
 DR WPI; 97-233185/21.
 PT DNA constructs contg. truncated promoter sequence - for
 fibre-specific gene expression in cotton plants
 PS Claim 1; Column 61-64; 48pp; English.

AC T30250;
 DT 16-DEC-1996 (first entry)
 DE Cotton fibre clone SKS1E6-H6-RI-derived promoter contg. sequence.
 KW Cotton fibre; Promoter; differential screening; leaf; ovule; root;
 OS flower; PCR; polymerase chain reaction; homology; transgenic plant; ds.
 Gossypium barbadense.
 KEY
 F1 promoter
 L1 .250
 FT /*tag= a
 FT /*note= "contains promoter sequence (claimed)"
 FT misc_feature
 FT 251 .. 321
 FT /*tag= b
 FT /*note= "5' non-coding sequence"
 FT 322 .. 1547
 FT /*tag= c
 FT /*note= "contains a 583 bp intron"
 FT 1548 .. 1984
 FT /*tag= d
 FT /*note= "3' non-coding sequence"
 FT US5521078-A.
 PN 28-MAY-1996.
 PD 04-OCT-1988; 253243.
 FT misc_feature
 FT 21-NOV-1990; US-17239.
 PR 18-MAY-1992; US-885970.
 PR 19-OCT-1994; US-298829.
 PA (CETU) AGRACETUS INC.
 PI John M;
 DR WPI; 96-267794/27.
 PT Isolation of fibre-specific cotton promoter sequences - using
 selected DNA probes to screen genomic DNA fragments, for production
 of cotton fibres with improved characteristics
 PS Claim 1; Column 59-62; 46pp; English.
 CC Cotton fibre cell-specific promoter sequences were isolated by
 differential screening of a cotton plant cDNA library. Of 4788 clones
 present in the leaf were isolated. These were screened with leaf cDNAs, 800 clones not
 ovule, root and flower mRNAs and resulted in 79 clones isolated. PCR
 analysis was then used to remove cross-hybridising clones. This resulted
 in the isolation of 18 cDNA clones specifically expressed in cotton fibre

Query Match 6.2%; Score 36.6; DB 1; Length 1985;
 Best Local Similarity 58.9%; Pred. No. 1..3;
 Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 TCATACCAATTAGACTCTCGCGTCGATCCCCCTGATACCGCTGTTCAGGCC 257
 DB 569 TCCACCCAGTTCTCCTCTCTGAACTCCACCTGCTTCCTCCGAACTCC 628

QY 258 AGCCAAAGCAGCGCGCCGGCTGGCGCATGGCTGAGCGCG 304
 DB 629 ACCTCCAGCTCTCCACCTGCTCCACCTCCACCGCTTCTCCAC 675

RESULT 15
 T86757/C
 ID T86757 standard; cDNA; 2415 BP.
 AC T86757;
 DT 15-DEC-1997 (first entry)
 DE cDNA of the M3/6 gene.
 KW murine; phosphatase; M3/6; dual specificity; Threonine; Tyrosine;
 KW activated protein kinase; MAP-K; cdc25 PTP; yeast;
 KW trinucleotide repeat; repeat expansion; neurodegenerative; proliferative;
 KW diagnosis; tumour; lung; brain; chromosomal deletion; ss.
 KW Mus sp.
 OS WO9106245-A1.
 PN

PD	20-FEB-1997	
PF	05-AUG-1996;	G01906.
PR	04-AUG-1995;	GB-016059.
PA	(MEDI-) MEDICAL RES COUNCIL.	
PI	Davies KE, Theodosiou A;	
PT	97-15423/14.	Murine mitogen activated protein (MAP) kinase phosphatase M3/6 - is a suspected dual specificity Thr-Tyr phosphatase, useful for diagnosing and treating neuro-degenerative or proliferative diseases

PT e.g. tumours
PS Claim 1; Fig 1; 51pp; English.
CC This cDNA is said to encode a murine phosphatase designated M3/6 (see
CC also T86758). M3/6 is a suspected dual specificity Threonine/Tyrosine
CC phosphatase, capable of inactivating mitogen activated protein (MAP)
CC kinase. The M3/6 protein product shows high homology to the cd25/PTP
CC of yeast at residues 29-49 and 117-116. The gene also contains a complex
CC triplet distal to the catalytic domain which is translated into the
CC protein. This domain comprises of 4 serine residues which in turn
CC is followed by a further run comprising 23 serine residues which is
CC interrupted near the N-terminal section by a single asparagine. This
CC makes the phosphatase gene a candidate for a human disease caused by
CC repeat expansion or mutation. M3/6 is expressed highly in the brain and
CC may have utility in investigating signal transduction mechanisms in brain
CC and muscle. The M3/6 and Hb5 (a human homologue) genes may be
CC responsible, if mutated, for various neurodegenerative or proliferative
CC diseases, and may therefore be used for the diagnosis of such diseases,
CC e.g. tumours, especially lung or brain tumours, associated with deletion
CC of the chromosomal region l1p15.5. The polypeptides can be used to screen
CC for inhibitors to treat these diseases.

Sequence 2415 BP; 491 A; 764 C; 678 G; 465 T;

Query Match 6.1%: Score 36.4: DB 1: Lenath 2415:

	Best Local Similarity	Pred.	No. 1-6;	Indels	Gaps	0;
Qy	186	GGGAGTACAGCTATACCAATAAGACTTCATGGCTGCATCCCTGCATCCCCC	245			
Db	1869	GGTCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT				1810
Qy	246	TGTTTCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCC	287			
Db	1809	TGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG				1768

Search completed: April 2, 2000, 21:55:09
Job time: 11056 sec

THIS PAGE IS MARKED (USPTO)